VisTrails: A Python-based Scientific Workflow and Provenance System

David Koop

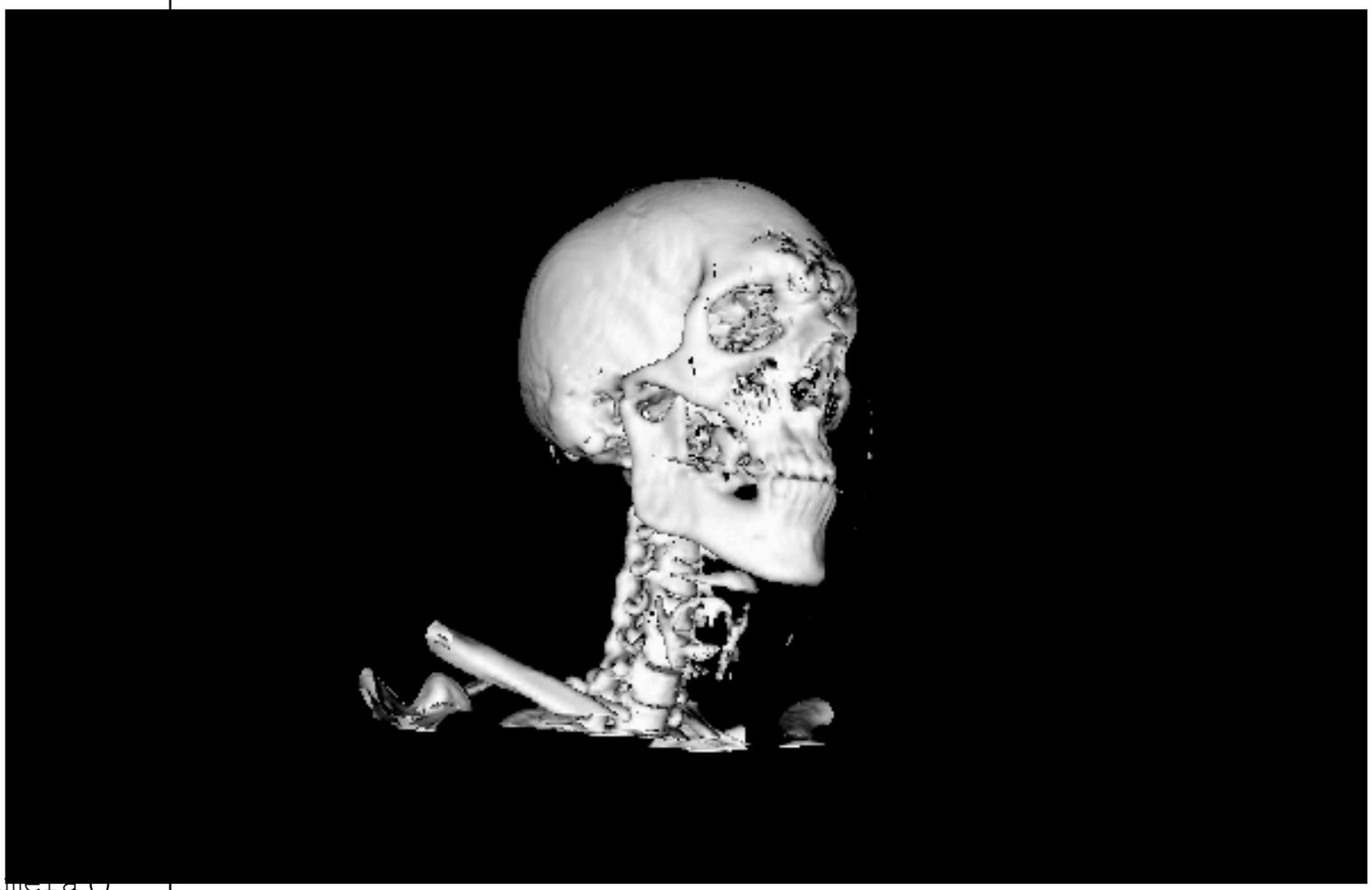
University of Utah and VisTrails, Inc.

VisTrails - Visual Workflow Programming

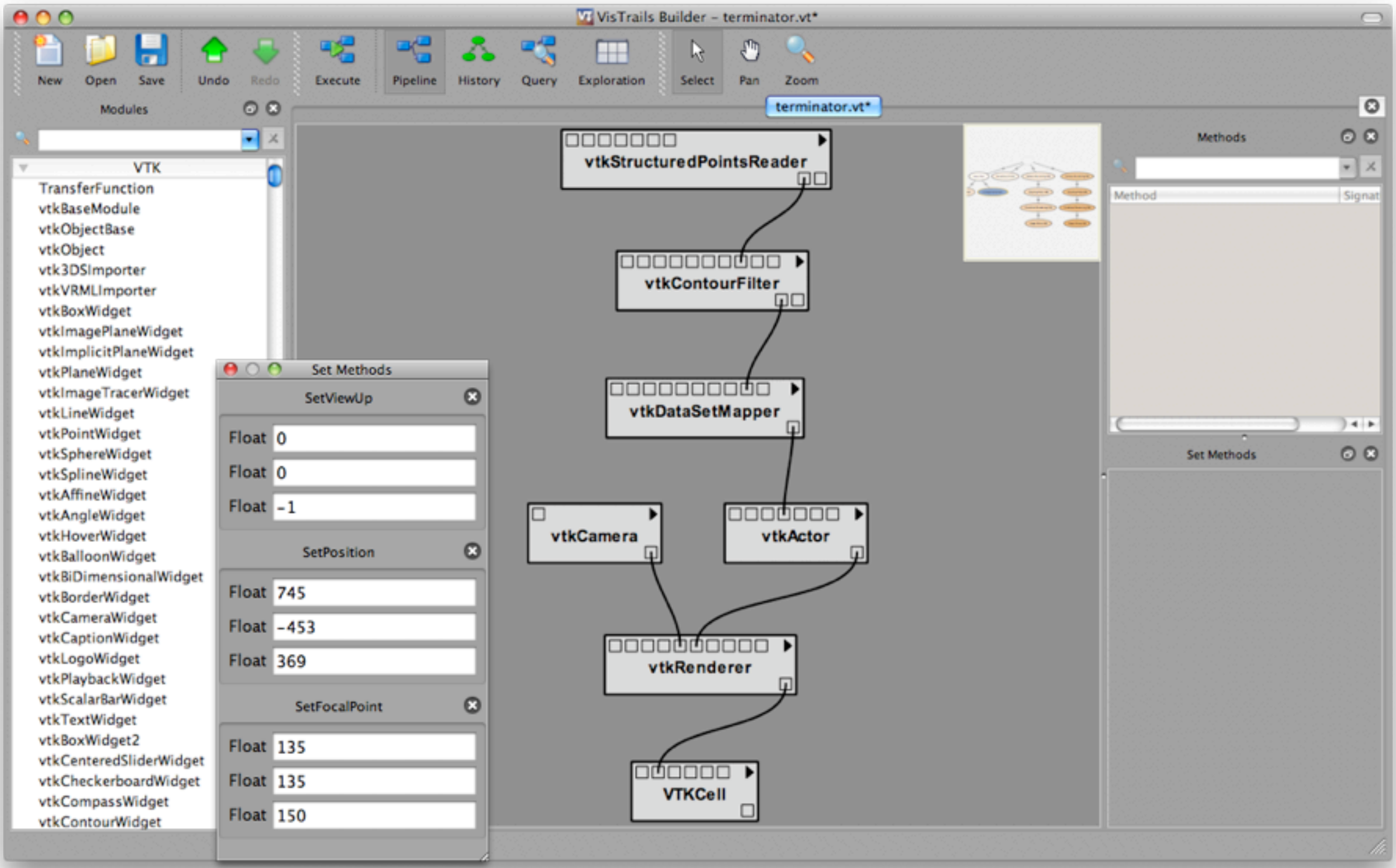
```
import vtk

data = vtk.vtkStructuredPointsReader()
data.SetFileName("../examples/data/head.120.vtk")

contour = vtk.vtkContourFilter()
contour.SetInput(data.GetOutput())
contour.SetValue(0, 67)

mapper = vtk.vtkPolyDataMapper()
mapper.SetInput(contour.GetOutput())
mapper.ScalarVisibilityOff()

actor = vtk.vtkActor()
actor.SetMapper(mapper)

cam = vtk.vtkCamera()
cam.SetViewUp(0,0,-1)
cam.SetPosition(745,-453,369)
cam.SetFocalPoint(135,135,150)
cam.ComputeViewPlaneNormal()

ren = vtk.vtkRenderer()
ren.AddActor(actor)
ren.SetActiveCamera(cam)
ren.ResetCamera()
renwin = vtk.vtkRenderWindow()
renwin.AddRenderer(ren)

style = vtk.vtkInteractorStyleTrackballCamera()
iren = vtk.vtkRenderWindowInteractor()
iren.SetRenderWindow(renwin)
iren.SetInteractorStyle(style)
iren.Initialize()
iren.Start()
```

VisTrails - Visual Workflow Programming

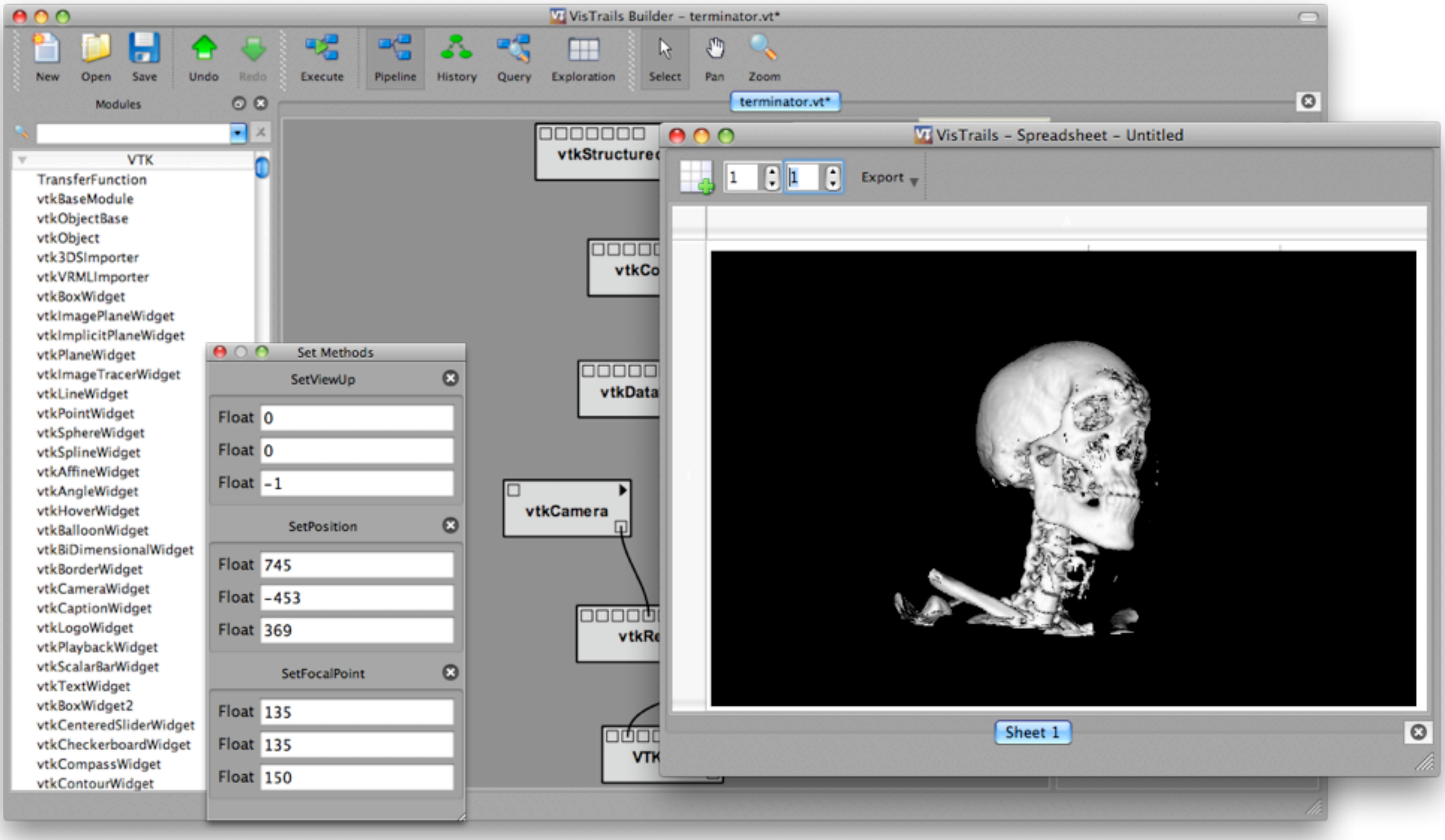
```
import vtk

data = vtk.vtkStructuredPointsReader()
data.SetFileName("../examples/data/head.120.vtk")

contour = vtk.vtkContourFilter()
contour.SetInput(data.GetOutput())
contour.SetValue(0, 67)

mapper = vtk.vtkPolyDataMapper()
mapper.SetInput(contour.GetOutput())
mapper.ScalarVisibilityOff()

actor = vtk.vtkActor()
actor.SetMapper(mapper)

cam = vtk.vtkCamera()
cam.SetViewUp(0,0,-1)
cam.SetPosition(745,-453,369)
cam.SetFocalPoint(135,135,150)
cam.ComputeViewPlaneNormal()

ren = vtk.vtkRenderer()
ren.AddActor(actor)
ren.SetActiveCamera(cam)
ren.ResetCamera()
renwin = vtk.vtkRenderWindow()
renwin.AddRenderer(ren)

style = vtk.vtkInteractorStyleTrackballCamera()
iren = vtk.vtkRenderWindowInteractor()
iren.SetRenderWindow(renwin)
iren.SetInteractorStyle(style)
iren.Initialize()
iren.Start()
```



VisTrails - Visual Workflow Programming



VisTrails - Visual Workflow Programming



Friday, February 19, 2010

Structure Adds Value

1. 

XML: Visual QuickStart Guide (2nd Edition) by Kevin Howard Goldberg (**Paperback** - Dec. 21, 2008)

Buy new: ~~$34.99~~ **$23.09**

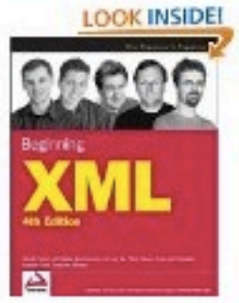
29 new from $19.99 14 used from $20.55

Get it by **Tuesday, Feb. 16** if you order in the next **2 hours** and choose one-day shipping.

(11) Prime

Excerpt - page 3: "... WRITING **XML** The **XML** specification defines how to write a document in ..."
Surprise me! See a random page in this book.

Books: See all 7,155 items

2. 

Beginning XML, 4th Edition (Programmer to Programmer) by David Hunter, Jeff Rafter, Joe Fawcett, and Eric van der Vlist (**Paperback** - May 21, 2007)

Buy new: ~~$39.99~~ **$26.39**

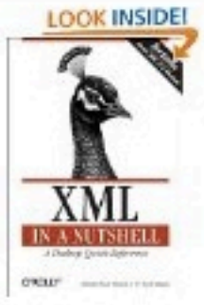
32 new from $25.02 19 used from $17.00

Get it by **Tuesday, Feb. 16** if you order in the next **1 hour** and choose one-day shipping.

(58) Prime

Excerpt - Front Matter: "... Beginning **XML** 4th Edition ..."
Surprise me! See a random page in this book.

Books: See all 7,155 items

3. 

XML in a Nutshell, Third Edition by Elliotte Rusty Harold and W. Scott Means (**Paperback** - Sept. 23, 2004)

Buy new: ~~$39.95~~ **$26.37**

34 new from $19.68 17 used from $14.99

Get it by **Tuesday, Feb. 16** if you order in the next **1 hour** and choose one-day shipping.

(17) Prime

Excerpt - page 3: "... Introducing **XML XML**, the Extensible Markup Language, is a W3C-endorsed standard for ..."
Surprise me! See a random page in this book.

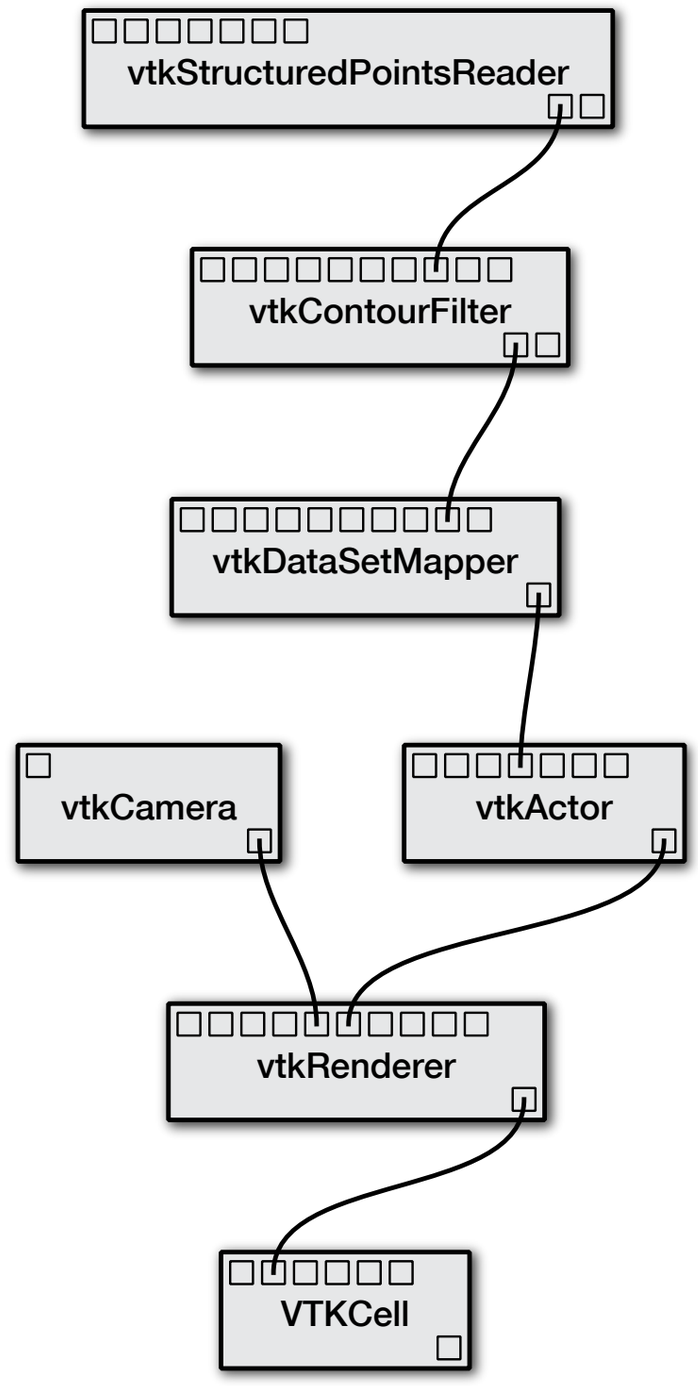
Books: See all 7,155 items

Structure Adds Value

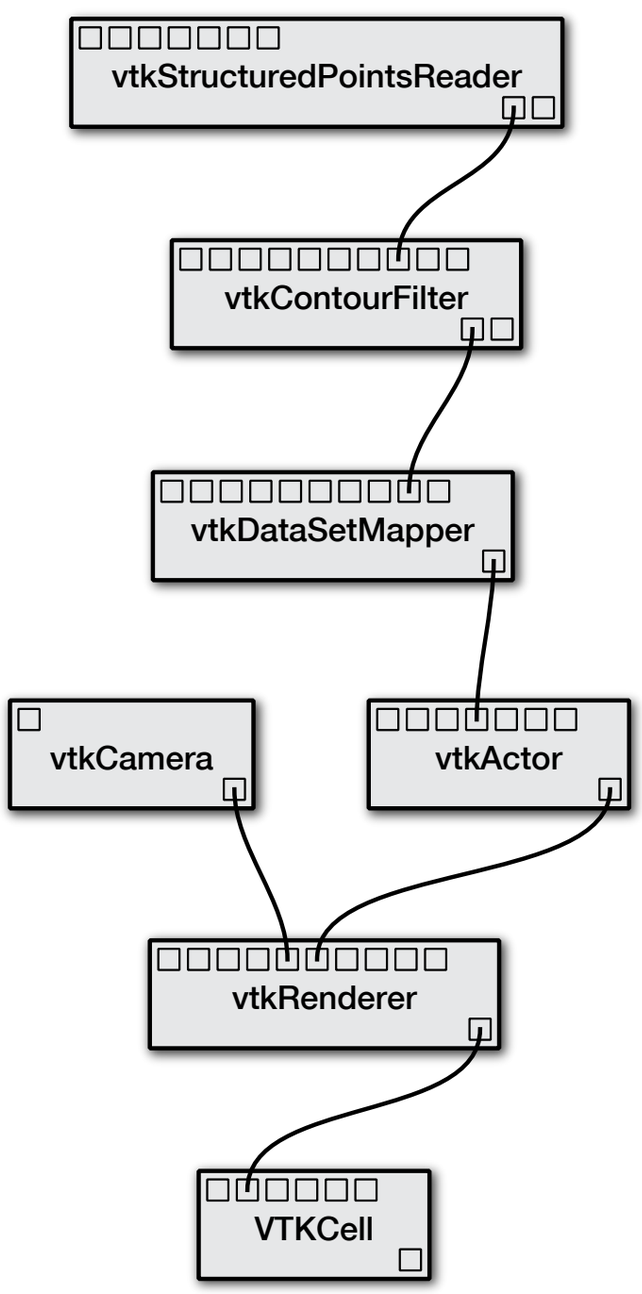

```
<book isbn="0321559673">
  <title>XML: Visual QuickStart Guide</title>
  <author>Kevin Howard Goldberg</author>
  <edition>2</edition>
  <price>23.09</price>
  <rating>5</rating>
</book>

<book isbn="0470114878">
  <title>Beginning XML</title>
  <author>David Hunter</author>
  <author>Jeff Rafter</author>
  <author>Joe Fawcett</author>
  <author>Eric van der Vlist</author>
  <edition>4</edition>
  <price>26.39</price>
  <rating>3.5</rating>
</book>

<book isbn="0596007647">
  <title>XML in a Nutshell</title>
  <author>Elliotte Rusty Harold</author>
  <author>W. Scott Means</author>
  <edition>3</edition>
  <price>26.37</price>
  <rating>4</rating>
</book>
```

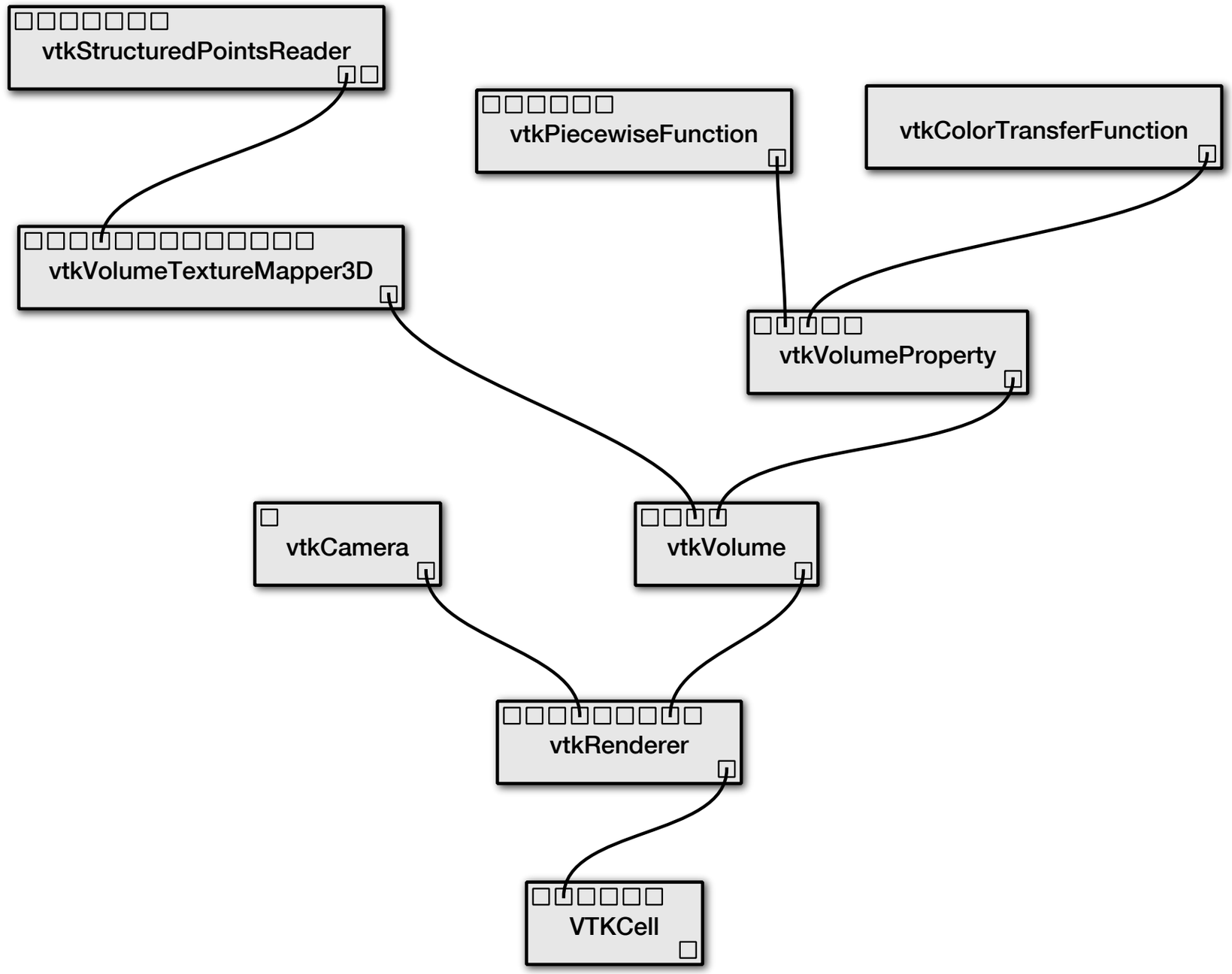
Structure Adds Value

```
import vtk

data = vtk.vtkStructuredPointsReader()
data.SetFileName(../examples/data/head.120.vtk)

contour = vtk.vtkContourFilter()
contour.SetInput(data.GetOutput())
contour.SetValue(0, 67)

mapper = vtk.vtkPolyDataMapper()
mapper.SetInput(contour.GetOutput())
mapper.ScalarVisibilityOff()

actor = vtk.vtkActor()
actor.SetMapper(mapper)

cam = vtk.vtkCamera()
cam.SetViewUp(0,0,-1)
cam.SetPosition(745,-453,369)
cam.SetFocalPoint(135,135,150)
cam.ComputeViewPlaneNormal()

ren = vtk.vtkRenderer()
ren.AddActor(actor)
ren.SetActiveCamera(cam)
ren.ResetCamera()
renwin = vtk.vtkRenderWindow()
renwin.AddRenderer(ren)

style = vtk.vtkInteractorStyleTrackballCamera()
iren = vtk.vtkRenderWindowInteractor()
iren.SetRenderWindow(renwin)
iren.SetInteractorStyle(style)
iren.Initialize()
iren.Start()
```

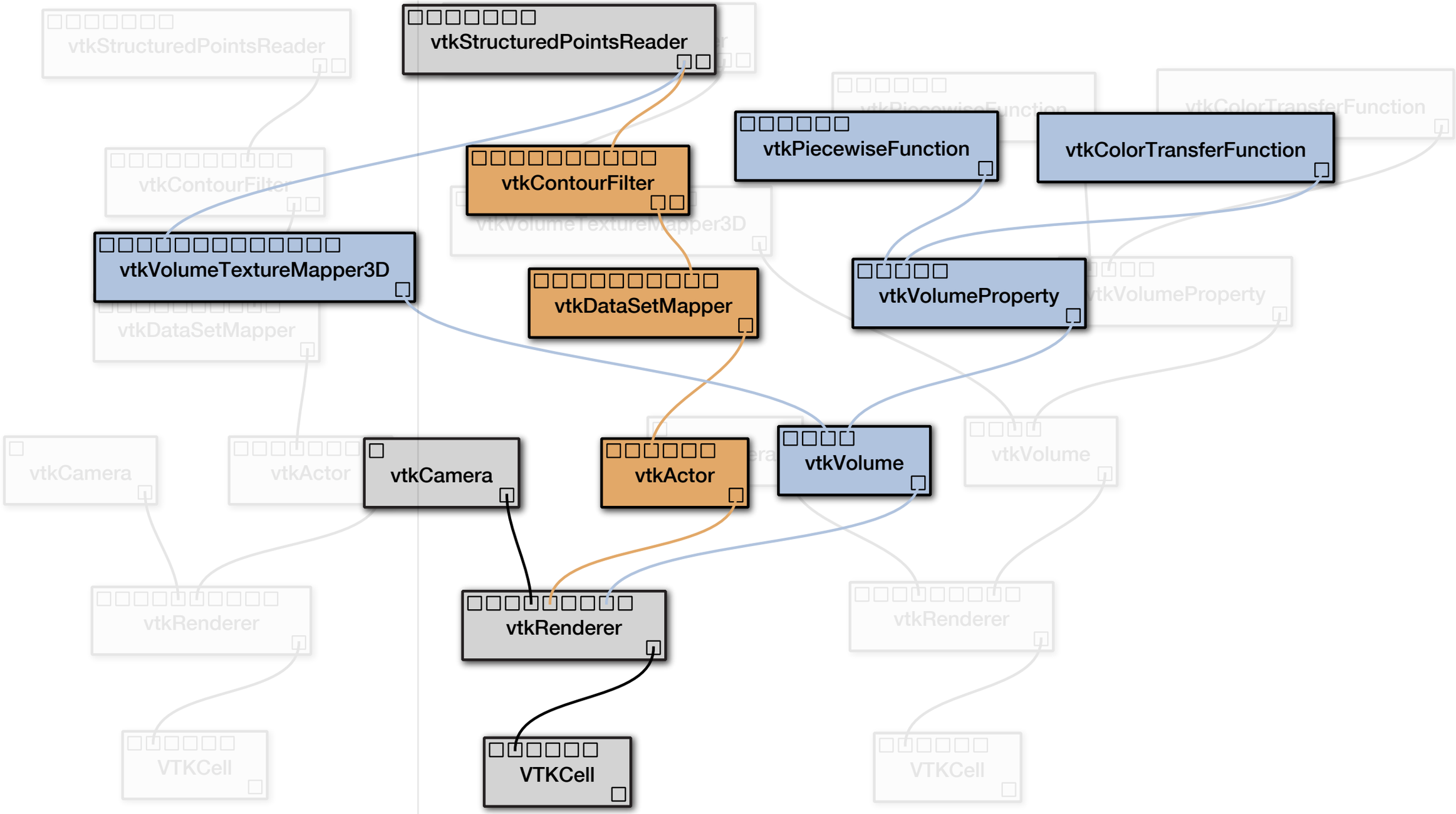


Workflow Differences

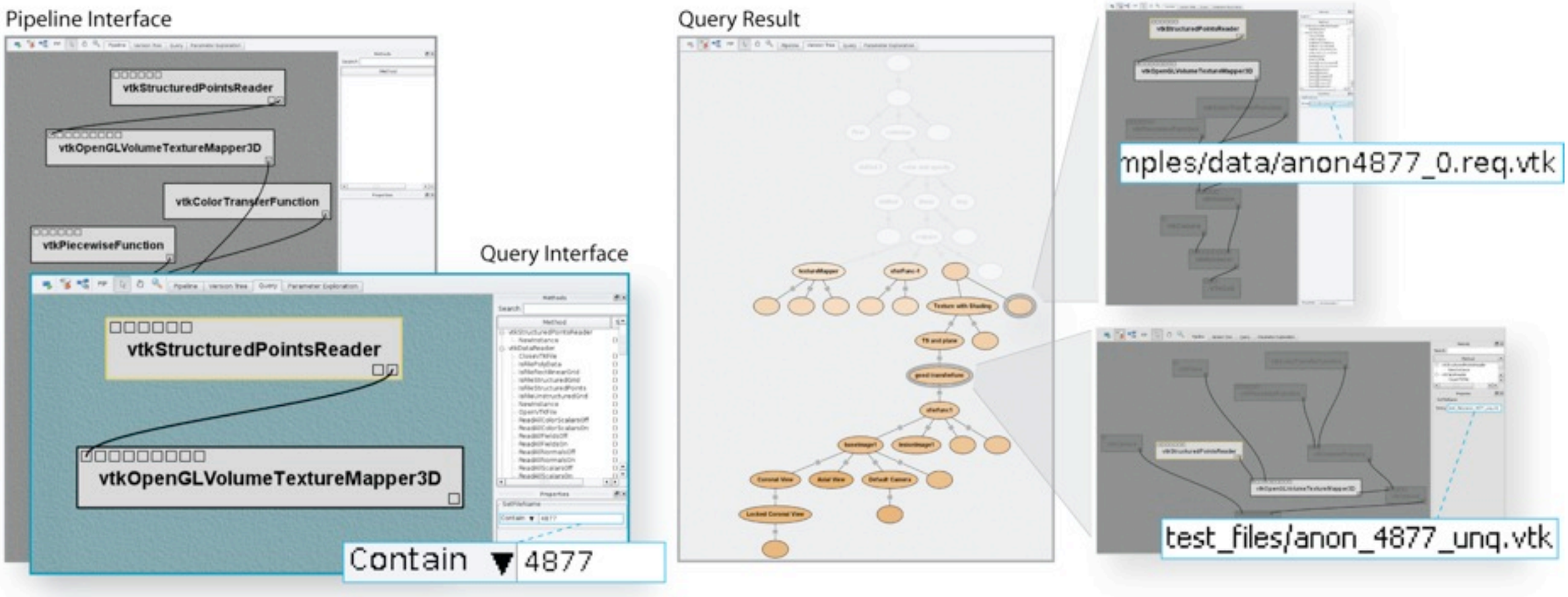




Workflow Differences

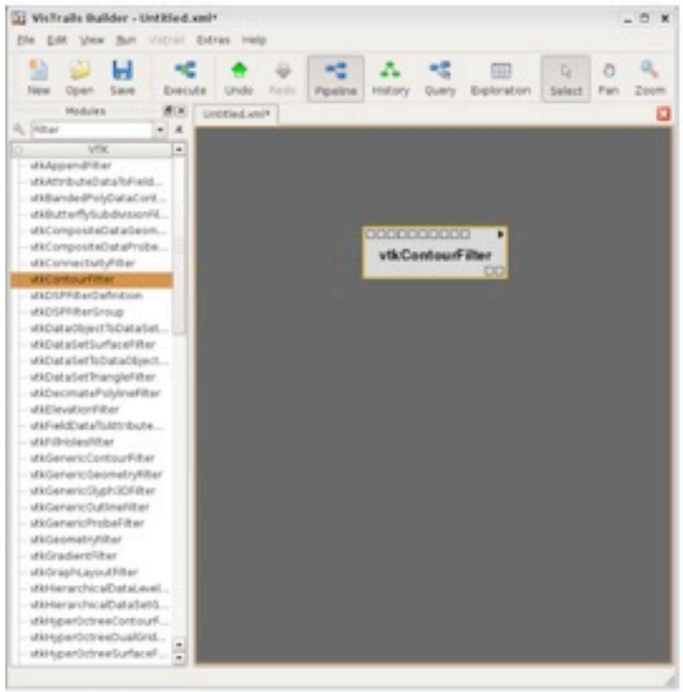


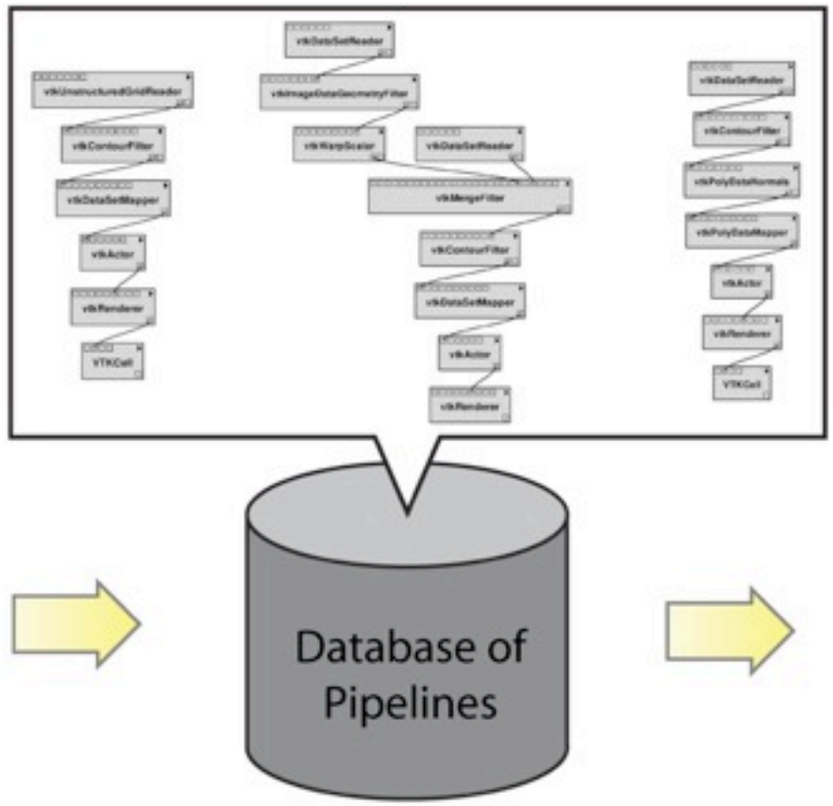
Workflow Query-by-example

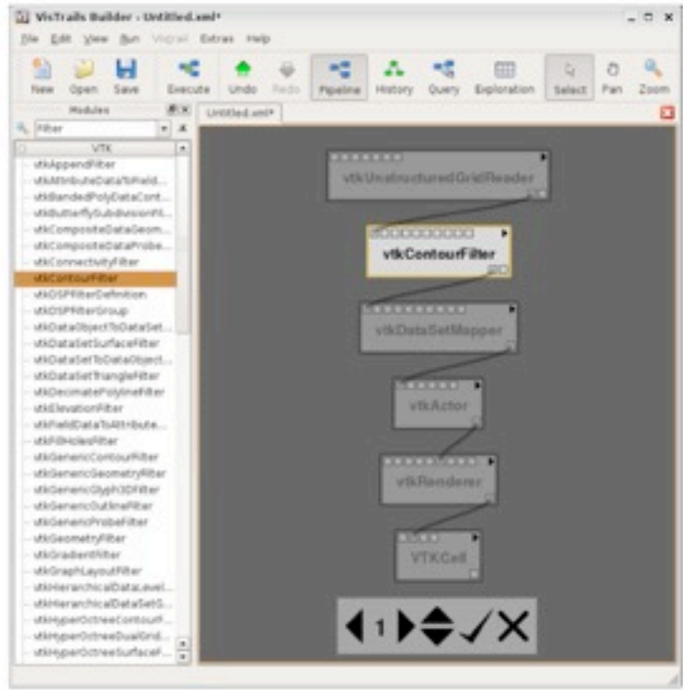


[IEEE Visualization 2007]

Workflow Completions







[IEEE Visualization 2008]

Provenance in Art

Rembrandt van Rijn
Dutch, 1606 - 1669
***Self-Portrait*, 1659**
oil on canvas
Andrew W. Mellon Collection
1937.1.72

Provenance

George, 3rd Duke of Montagu and 4th Earl of Cardigan [d. 1790], by 1767;[1] by inheritance to his daughter, Lady Elizabeth, wife of Henry, 3rd Duke of Buccleuch of Montagu House, London; John Charles, 7th Duke of Buccleuch; (P. & D. Colnaghi & Co., New York, 1928); (M. Knoedler & Co., New York); sold January 1929 to Andrew W. Mellon, Pittsburgh and Washington, D.C.; deeded 28 December 1934 to The A.W. Mellon Educational and Charitable Trust, Pittsburgh; gift 1937 to NGA.

[1] This early provenance is established by presence of a mezzotint after the portrait by R. Earlom (1743-1822), dated 1767. See John Charrington, *A Catalogue of the Mezzotints After, or Said to Be After, Rembrandt*, Cambridge, 1923, no. 49.

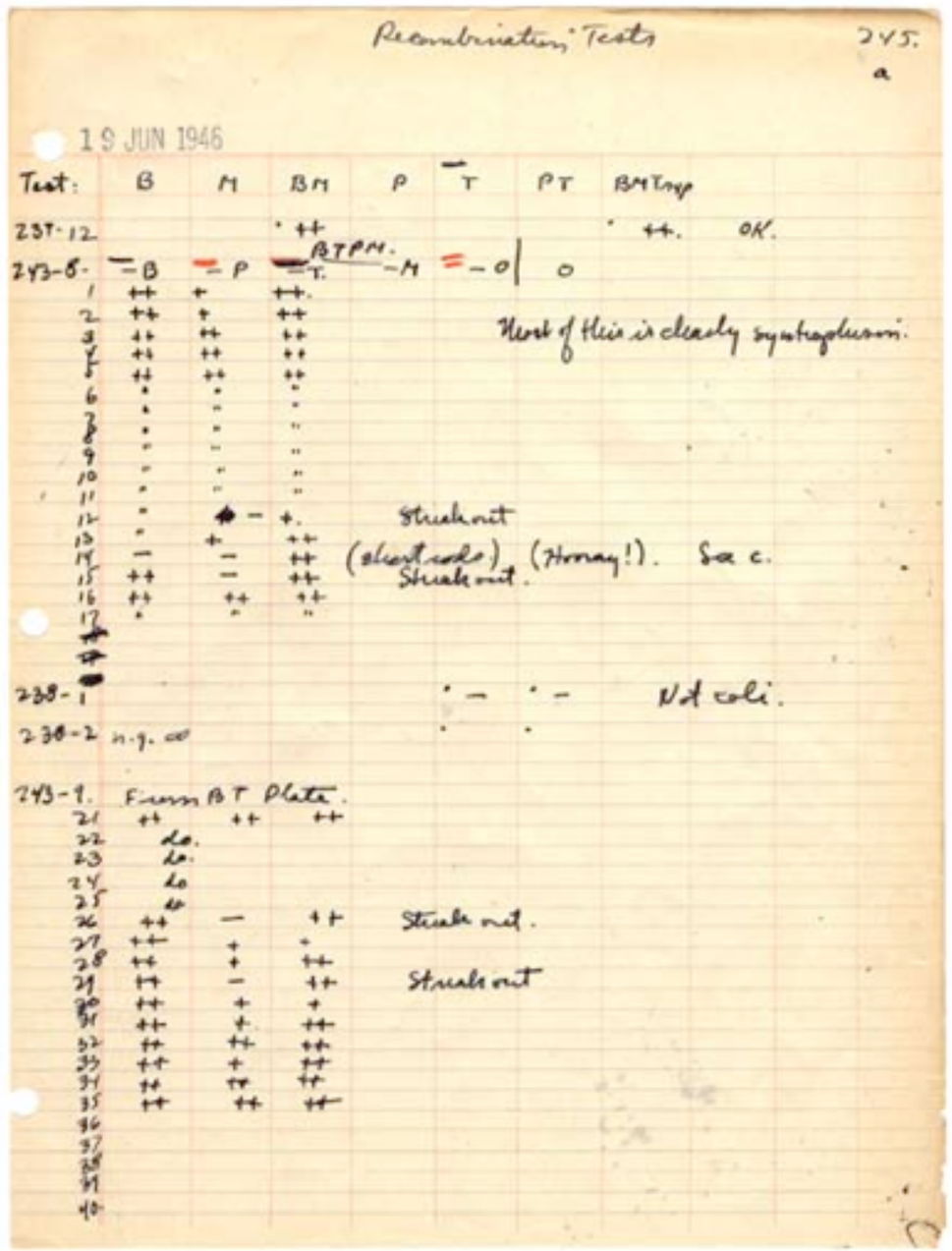
Associated Names

- Buccleuch, Henry, 3rd Duke of
- Buccleuch, John Charles, 7th Duke of
- Colnaghi & Co., Ltd., P. & D.
- Knoedler & Company, M.
- Mellon, Andrew W.
- Mellon Educational and Charitable Trust, The A.W.
- Montagu, and 4th Earl of Cardigan, George, 3rd Duke of

[National Gallery of Art]

Provenance in Science

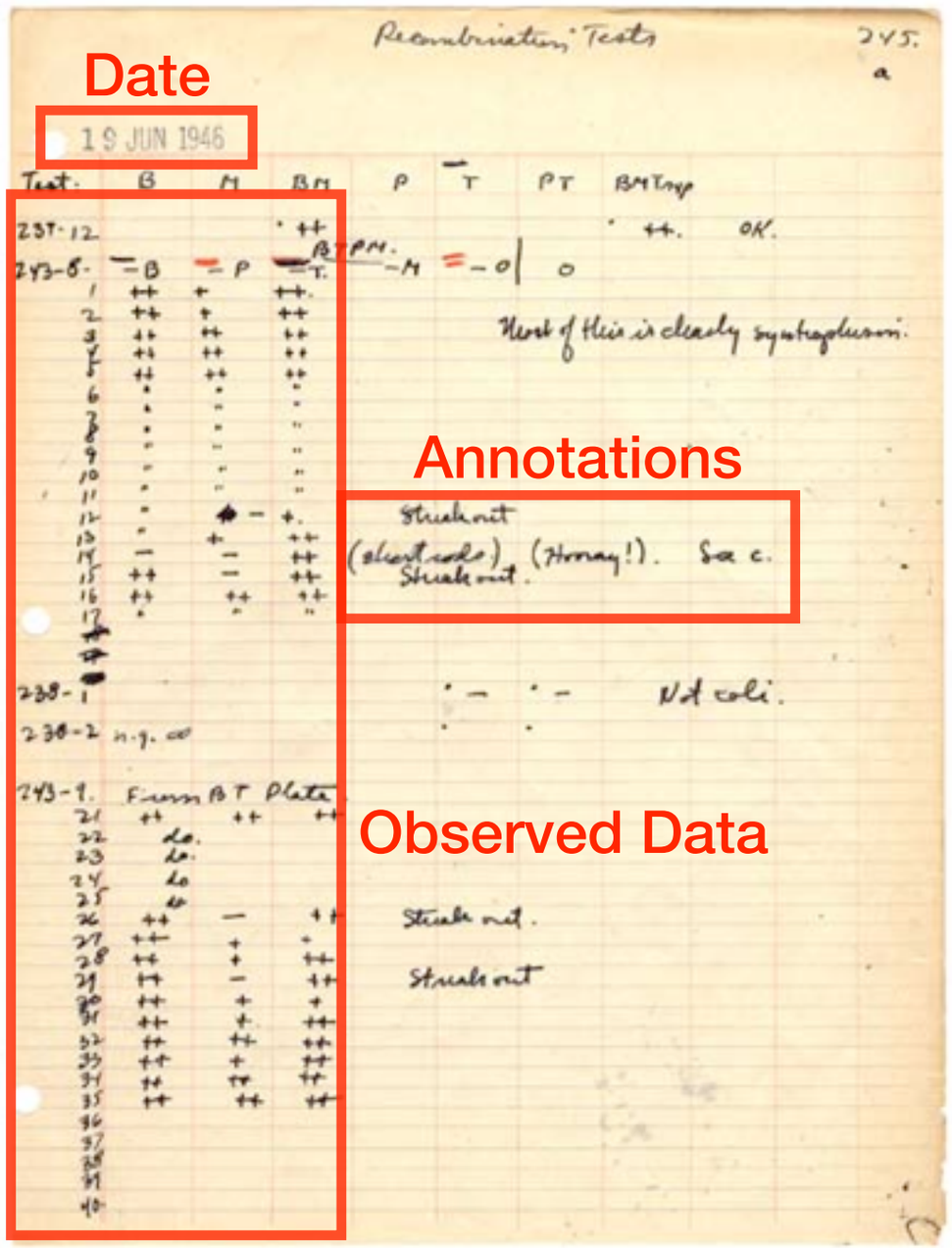
- **Provenance is as (or more) important as the result!**
- Old solution:
  - Lab notebooks
- New problems:
  - Large volumes of data
  - Complex analyses
  - Writing notes doesn't scale
- New solution:
  - Automated provenance capture with user-defined annotations



[DNA Recombination, Lederberg]

Provenance in Science

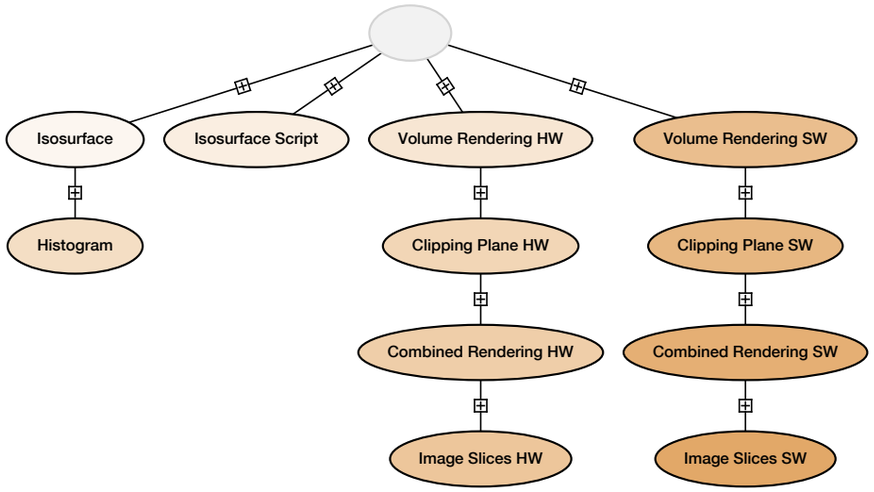
- **Provenance is as (or more) important as the result!**
- Old solution:
  - Lab notebooks
- New problems:
  - Large volumes of data
  - Complex analyses
  - Writing notes doesn't scale
- New solution:
  - Automated provenance capture with user-defined annotations



[DNA Recombination, Lederberg]

Provenance in VisTrails

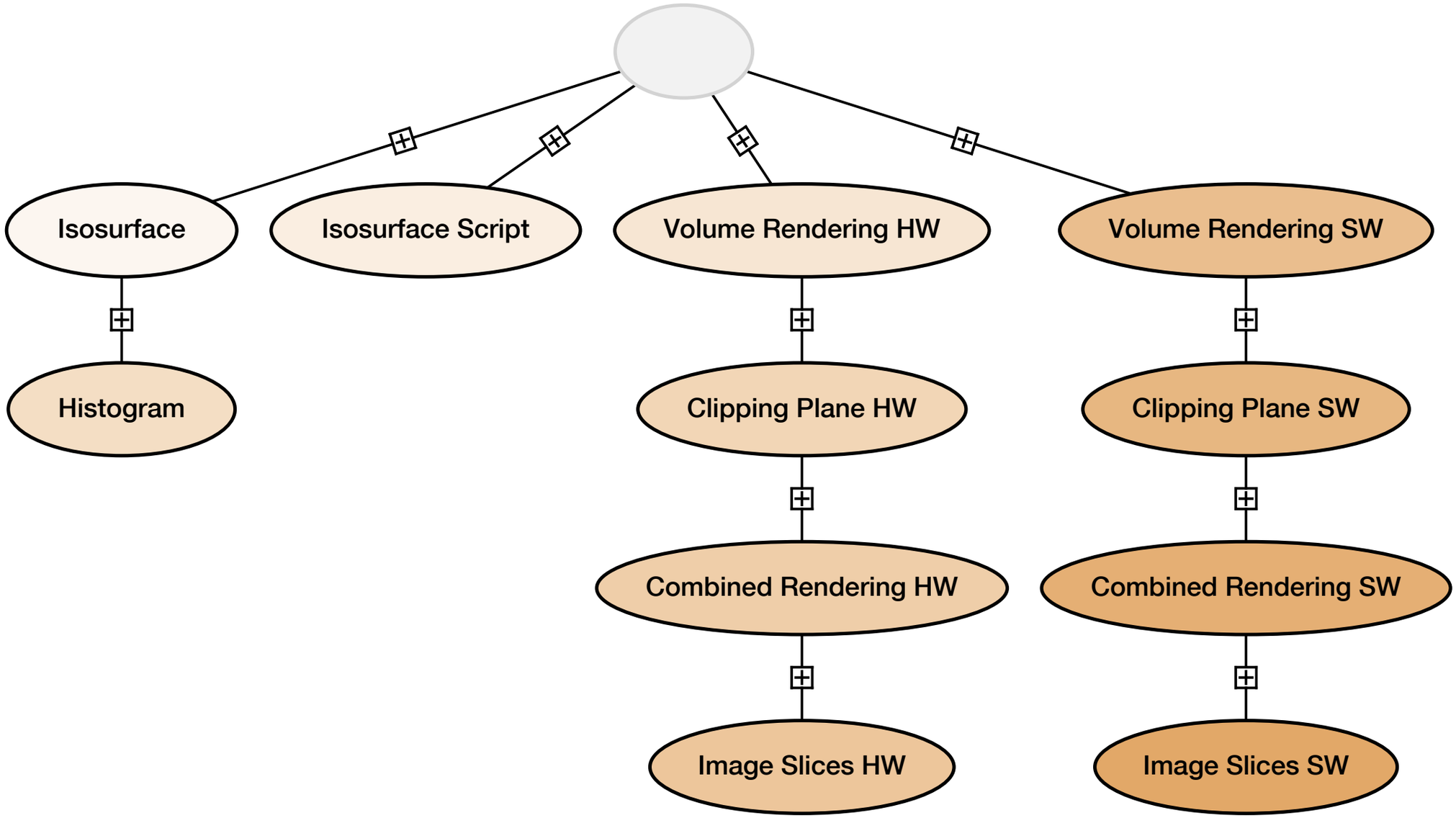
- Design Provenance (Version Tree)



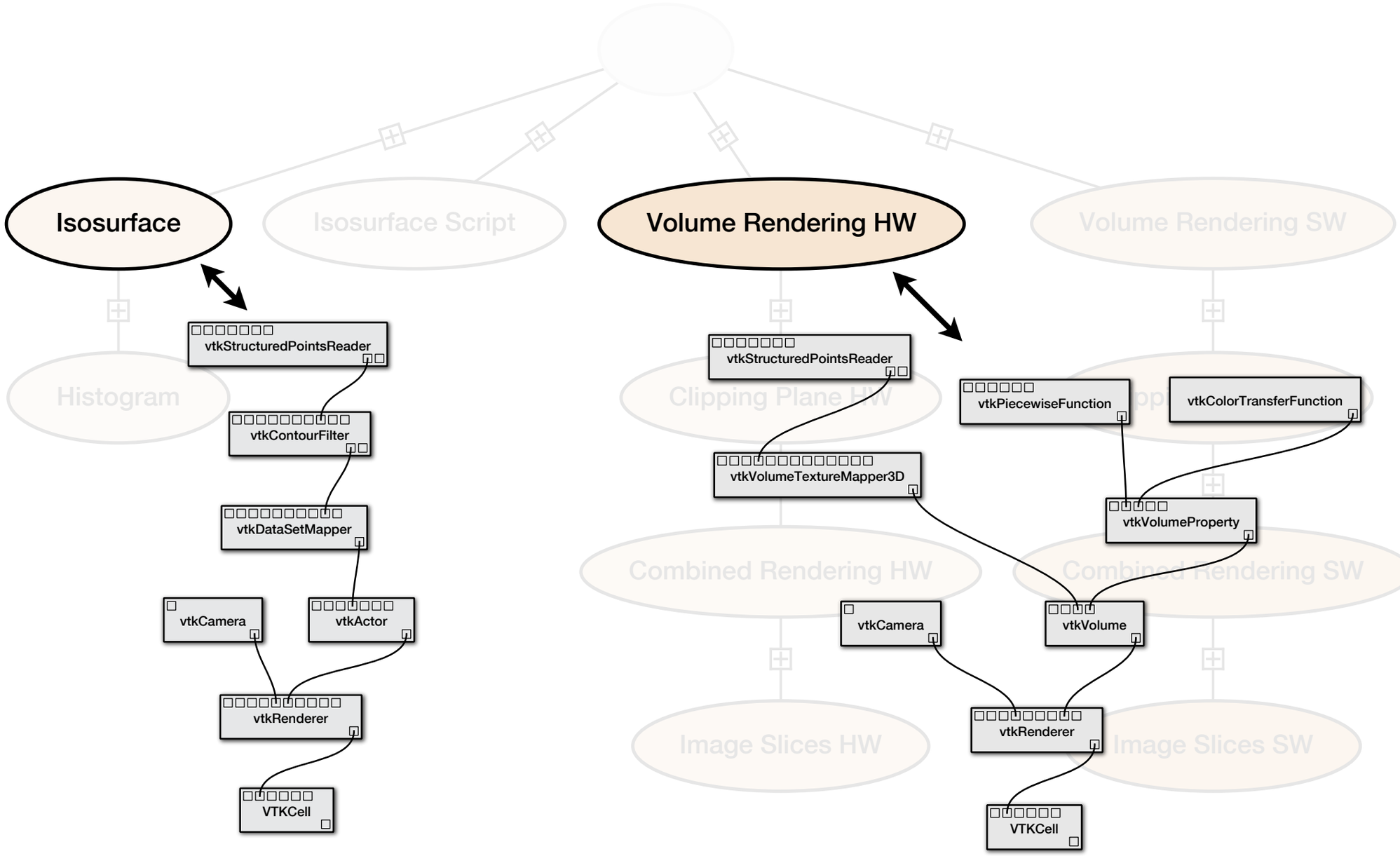
- Execution Provenance (Logging):

```
<module id="12" name="vtkDataSetReader" start_time="2010-02-19 11:01:05"
     end_time="2010-02-19 11:01:07">
  <annotation key="hash" value="c54bea63cb7d912a43ce"/>
</module>
<module id="13" name="vtkContourFilter" start_time="2010-02-19 11:01:07"
     end_time="2010-02-19 11:01:08"/>
<module id="15" name="vtkDataSetMapper" start_time="2010-02-19 11:01:09"
     end_time="2010-02-19 11:01:12"/>
```

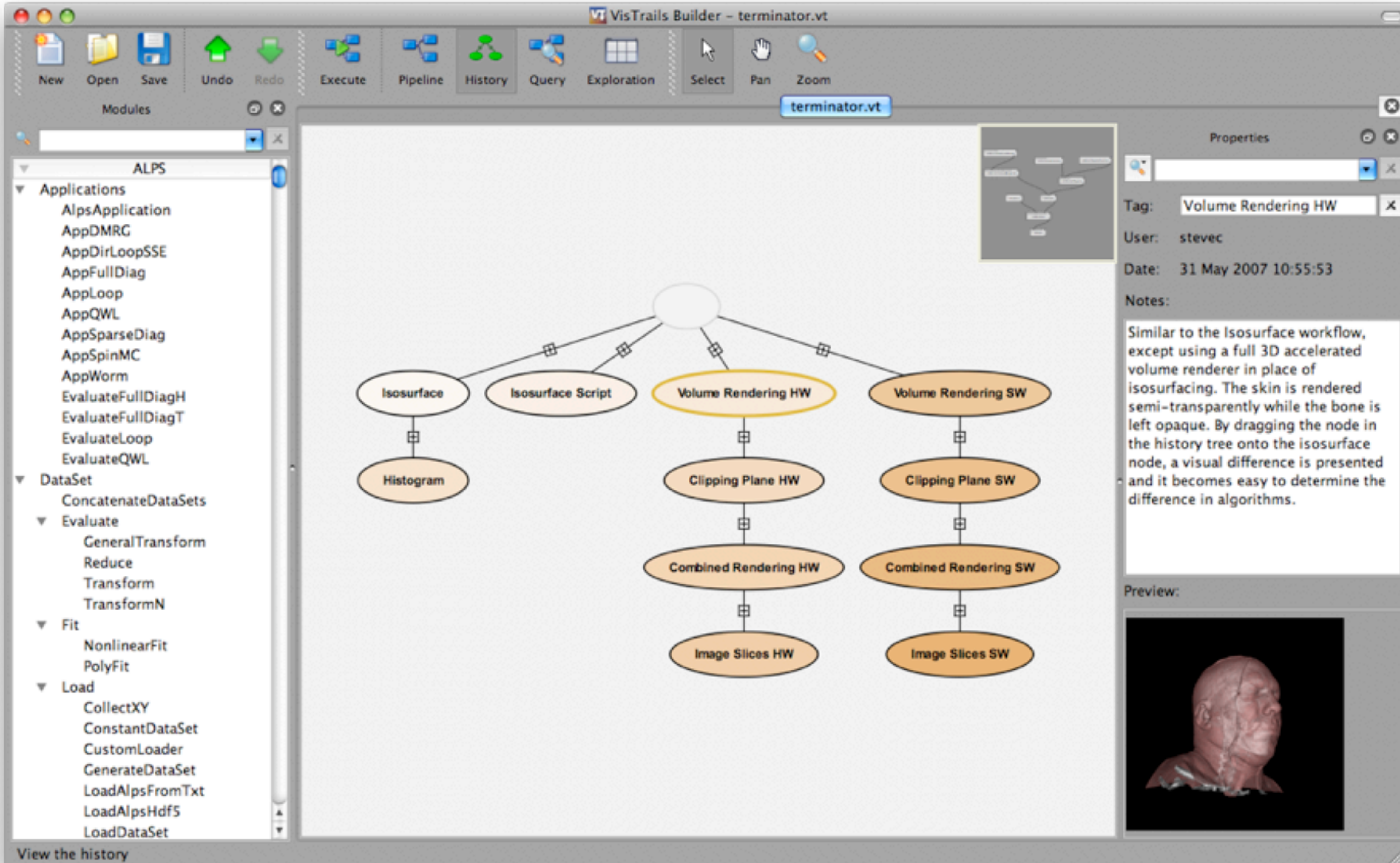
Design Provenance



Design Provenance



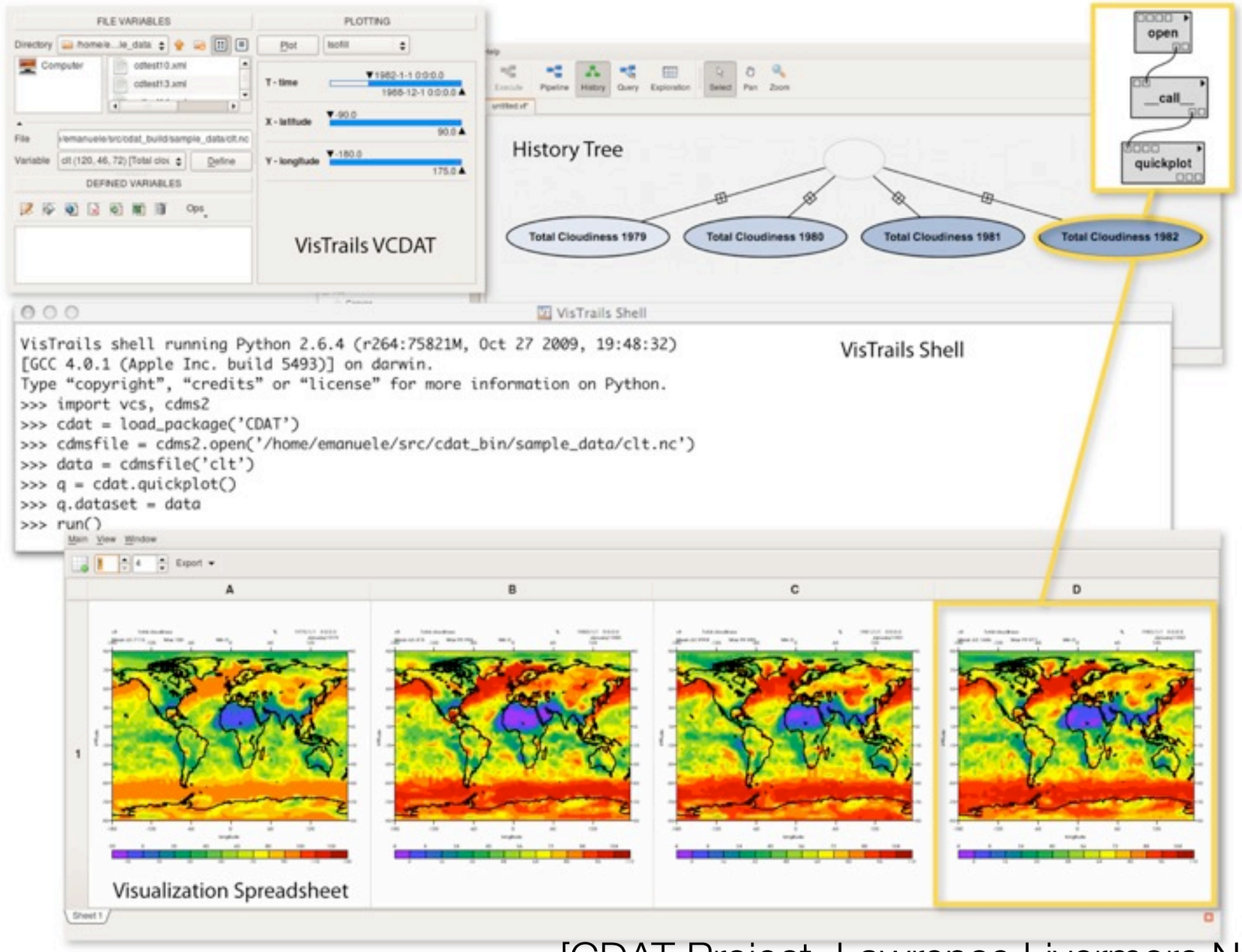
VisTrails - Design Provenance



Specifications

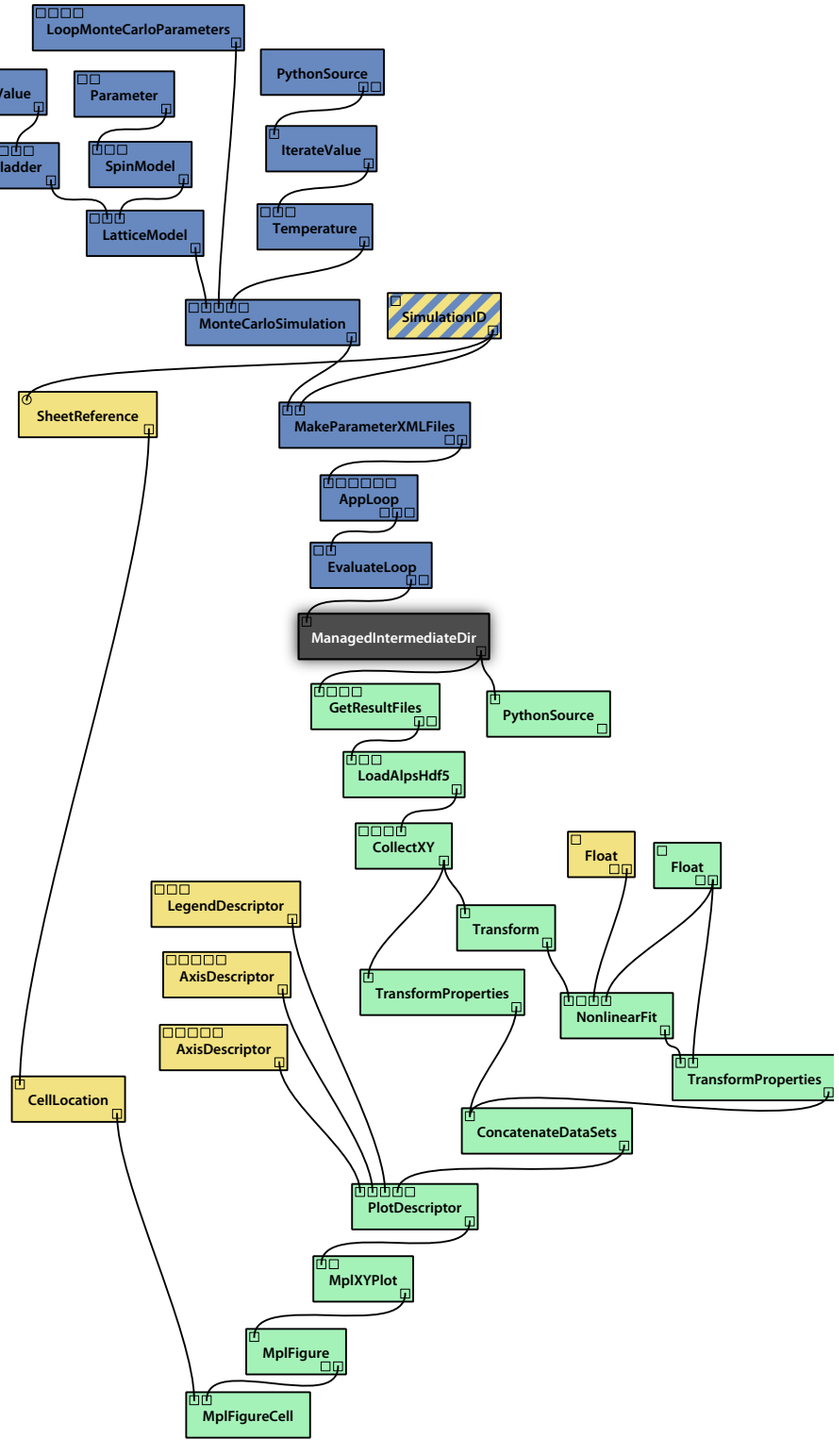
- Open-source, freely downloadable system (www.vistrails.org)
- Multi-platform: users on Mac, Linux, and Windows
- Python code and uses PyQt and Qt for the interface
- Over 13,000 downloads since 2007
- User's guide, wiki, and mailing list
- Many users in different disciplines and countries:
  - Visualizing environmental simulations (CMOP STC)
  - Simulation for solid, fluid and structural mechanics (Galileo Network, UFRJ Brazil)
  - Quantum physics simulations (ALPS, ETH Switzerland)
  - Climate analysis (CDAT)
  - Habitat modeling (USGS)
  - Open Wildland Fire Modeling (U. Colorado, NCAR)
  - High-energy physics (LEPP, Cornell)
  - Cosmology simulations (LANL)
  - Using tms for improving memory (Pyschiatry, U. Utah)
  - eBird (Cornell, NSF DataONE)
  - Astrophysical Systems (Tohline, LSU)
  - NIH NBCR (UCSD)
  - Pervasive Technology Labs (Heiland, Indiana University)
  - Linköping University
  - University of North Carolina, Chapel Hill
  - UTEP

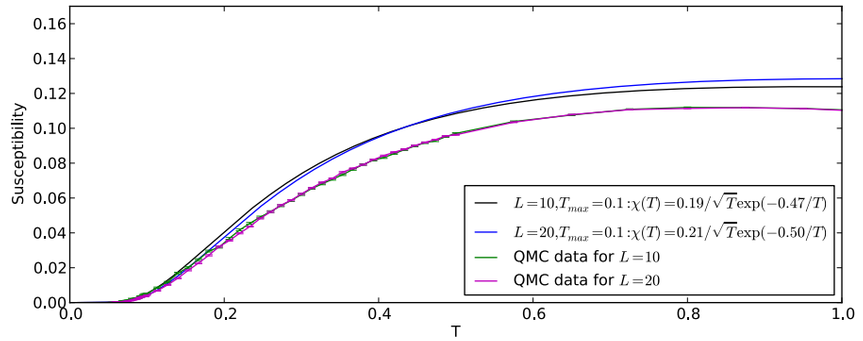
Climate Data Analysis

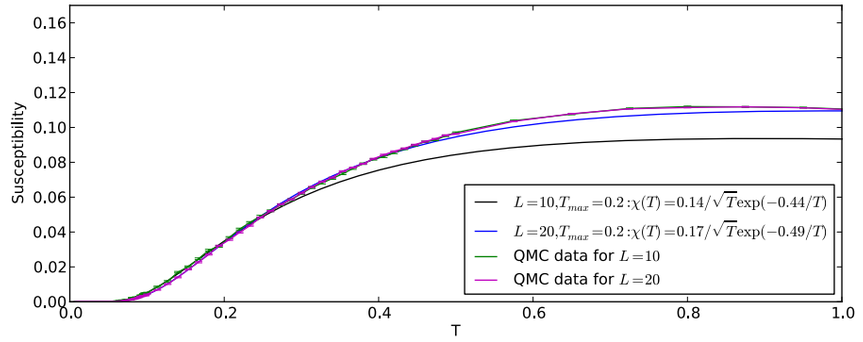


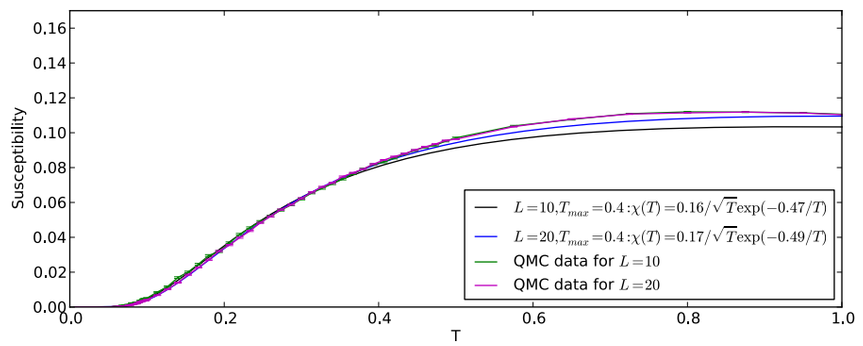
[CDAT Project, Lawrence Livermore National Lab]

Quantum Lattice Models



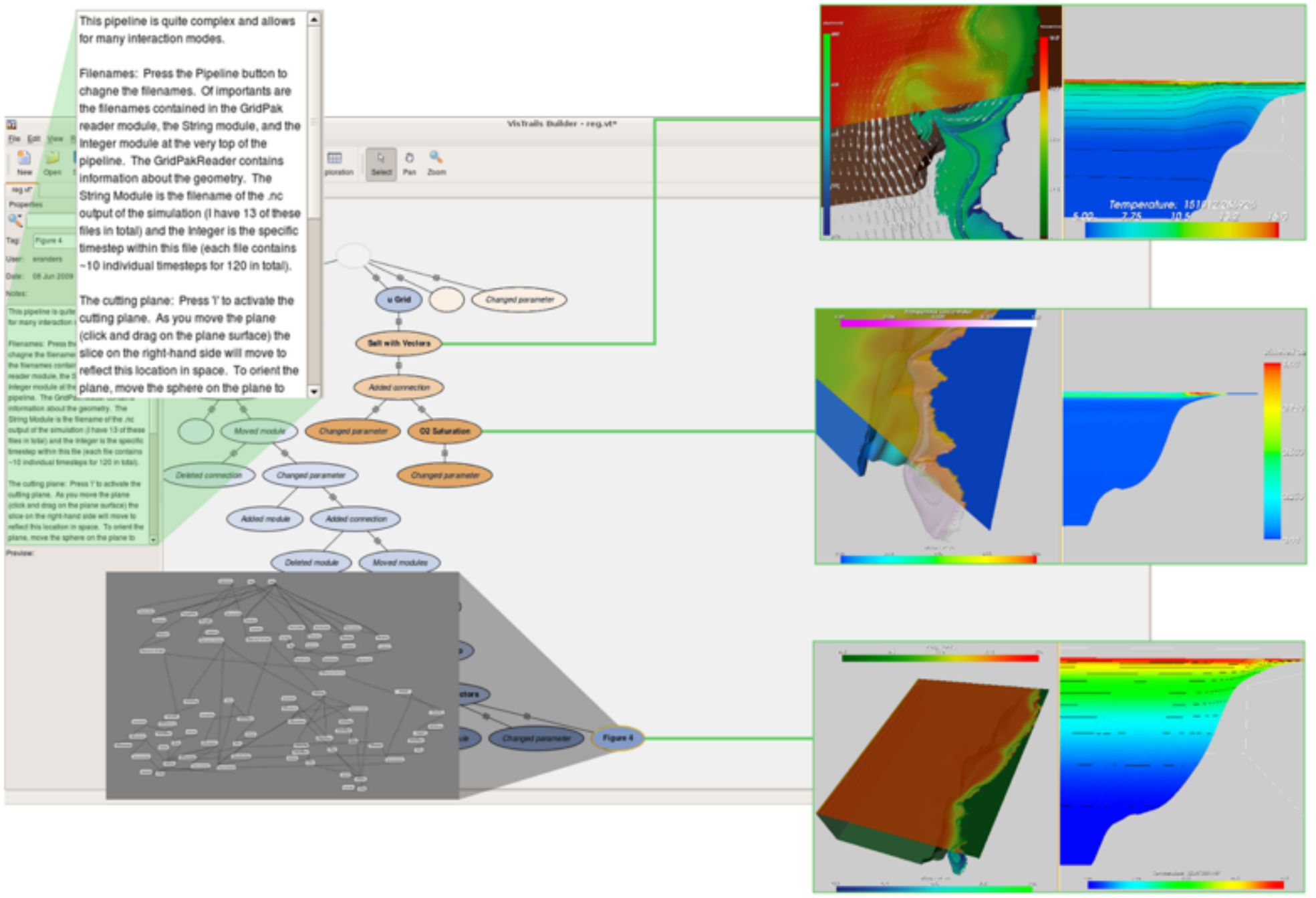






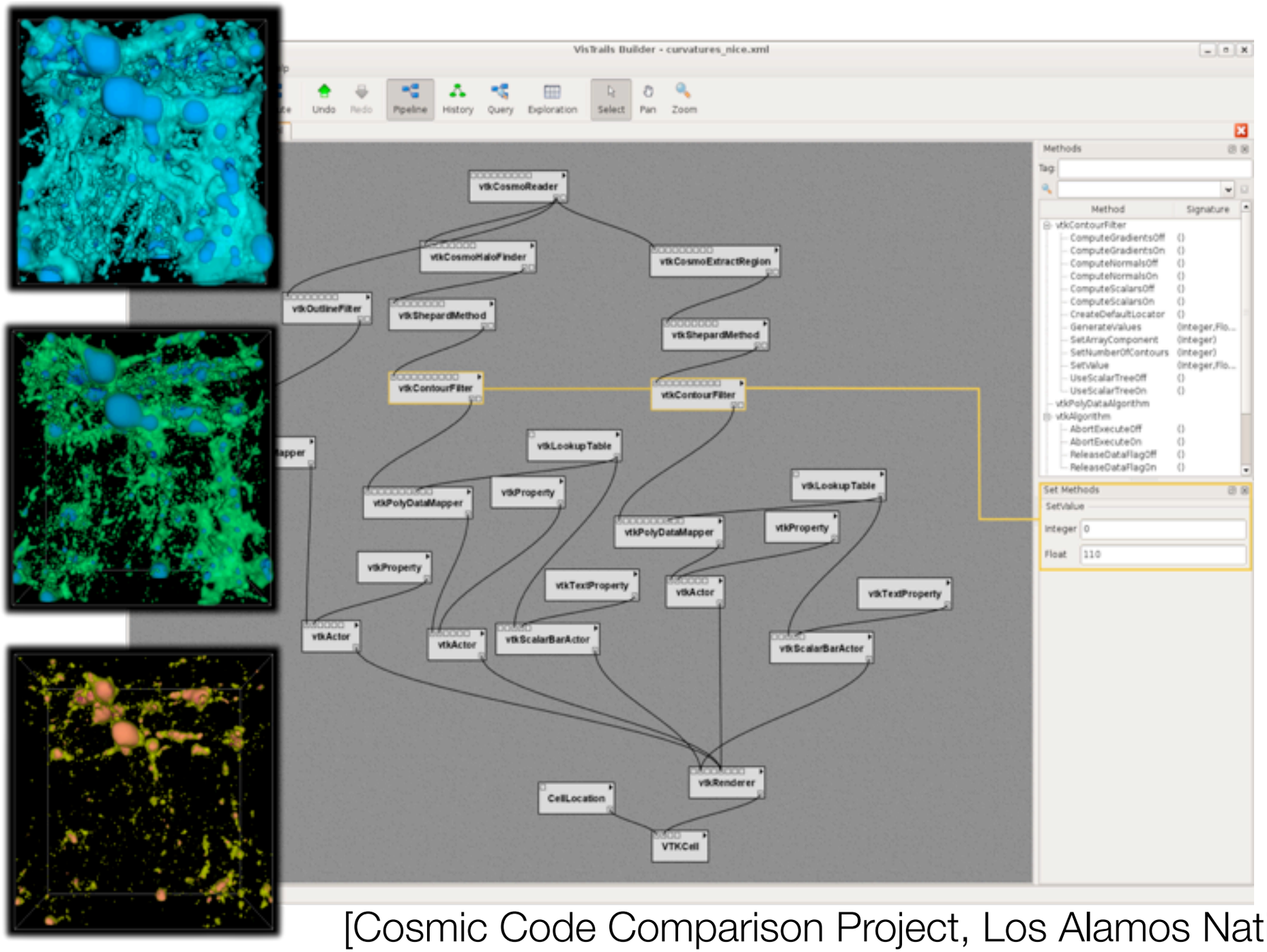
[ALPS Project, ETH-Zurich]

Coastal Margin Observation & Prediction



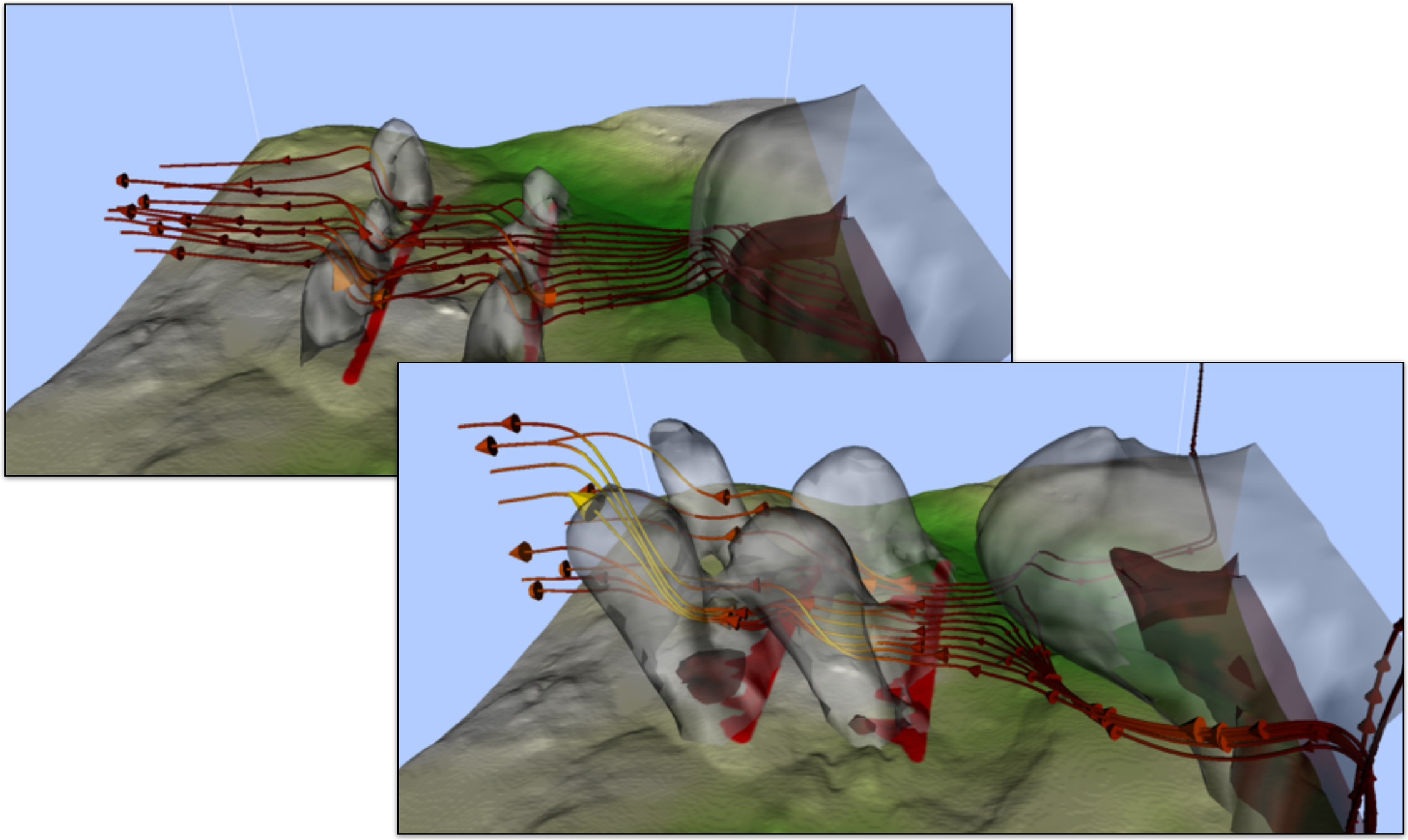
[NSF Science & Technology Center for Coastal Margin Observation & Prediction]

Comparing Cosmological Simulations



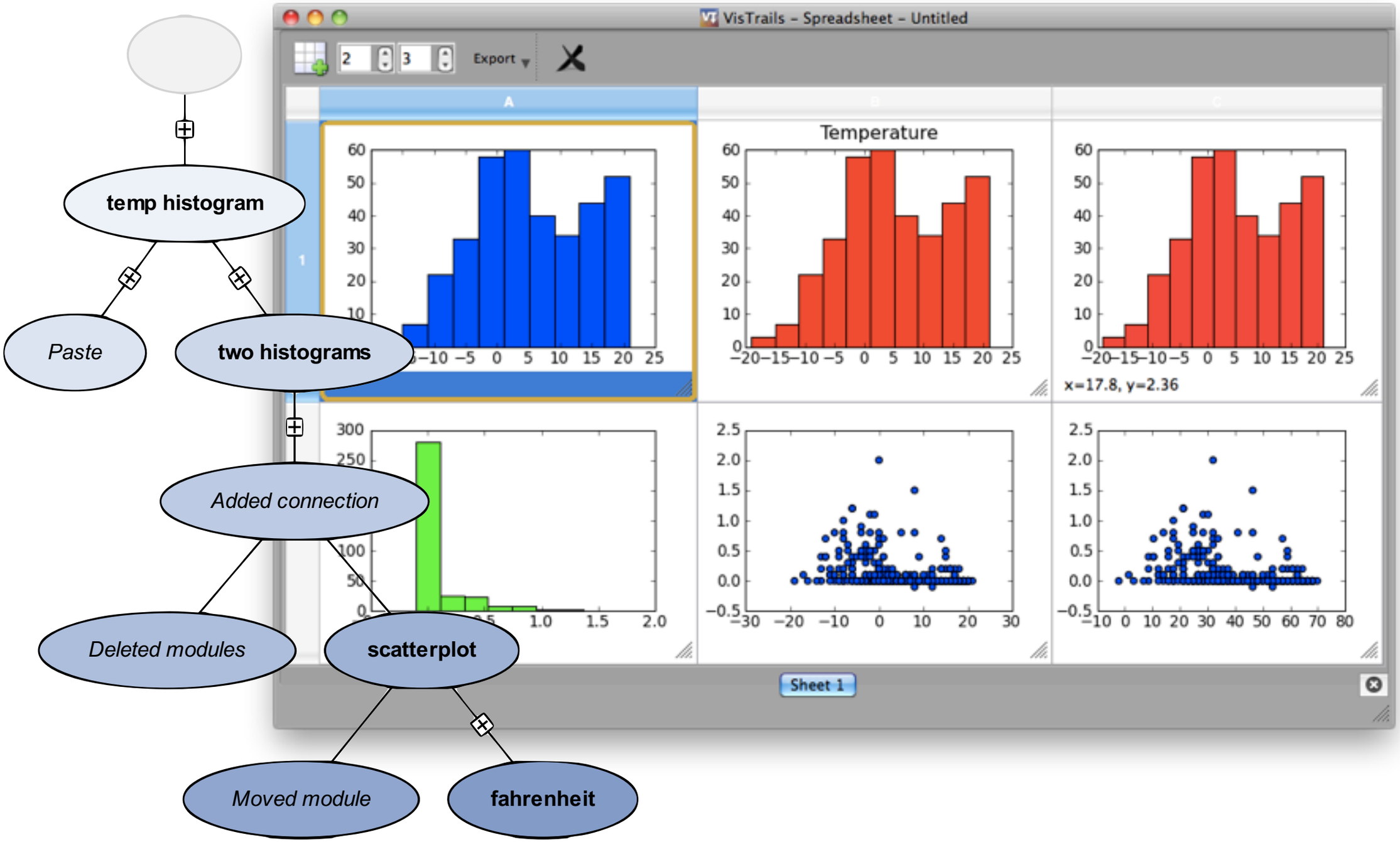
[Cosmic Code Comparison Project, Los Alamos National Lab]

Wildfire Prediction



[WRF-Fire Project, University of Colorado-Denver]

Demo


VisTrails - Spreadsheet - Untitled
Export
Temperature
x=17.8, y=2.36
Sheet 1
temp histogram
Paste
two histograms
Added connection
Deleted modules
scatterplot
Moved module
fahrenheit

Extensibility

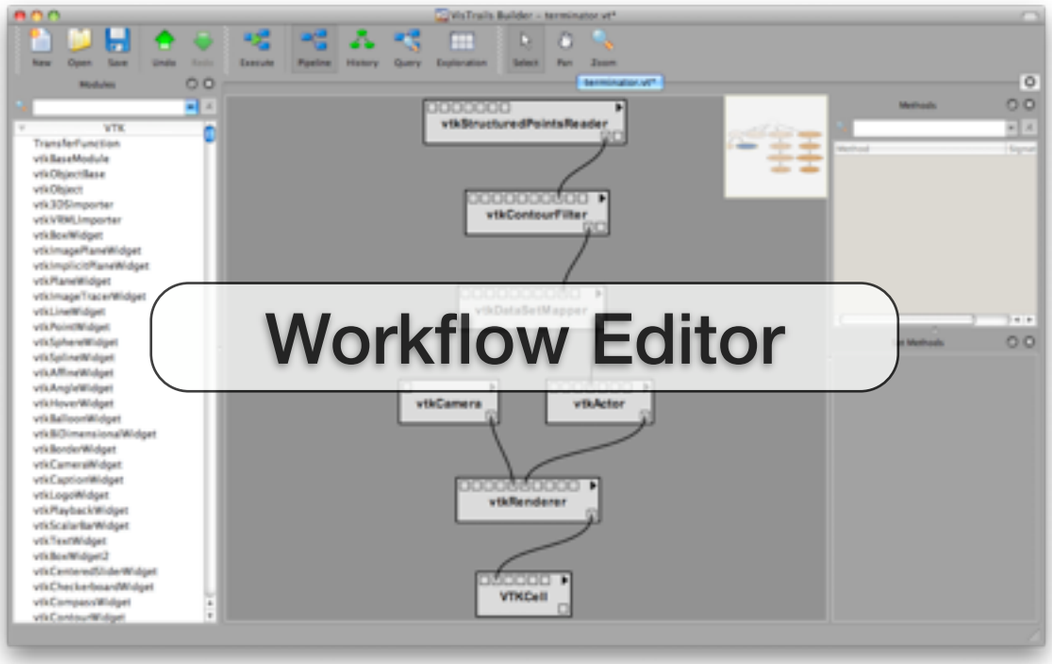
- Package infrastructure
- Wrap python libraries, command-line calls, or use other interfaces (jpype, rpy, etc.)
- Need to specify:
  1. Package identification information
  2. Module structures: input & output ports
  3. Compute method for each module

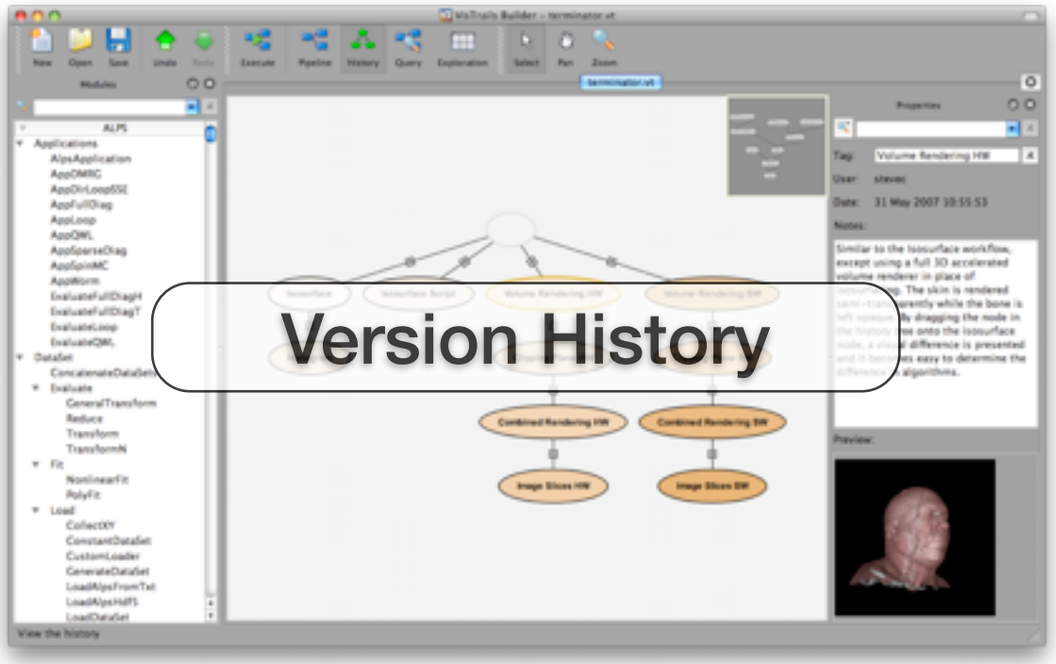
Extensibility Example

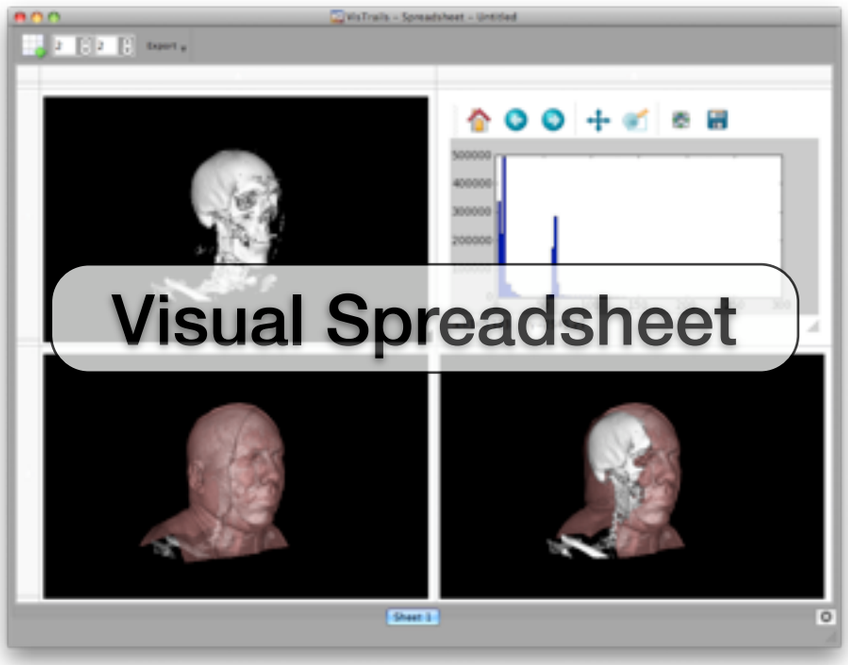
- seawater python package:
  - http://pypi.python.org/pypi/seawater/1.0.3

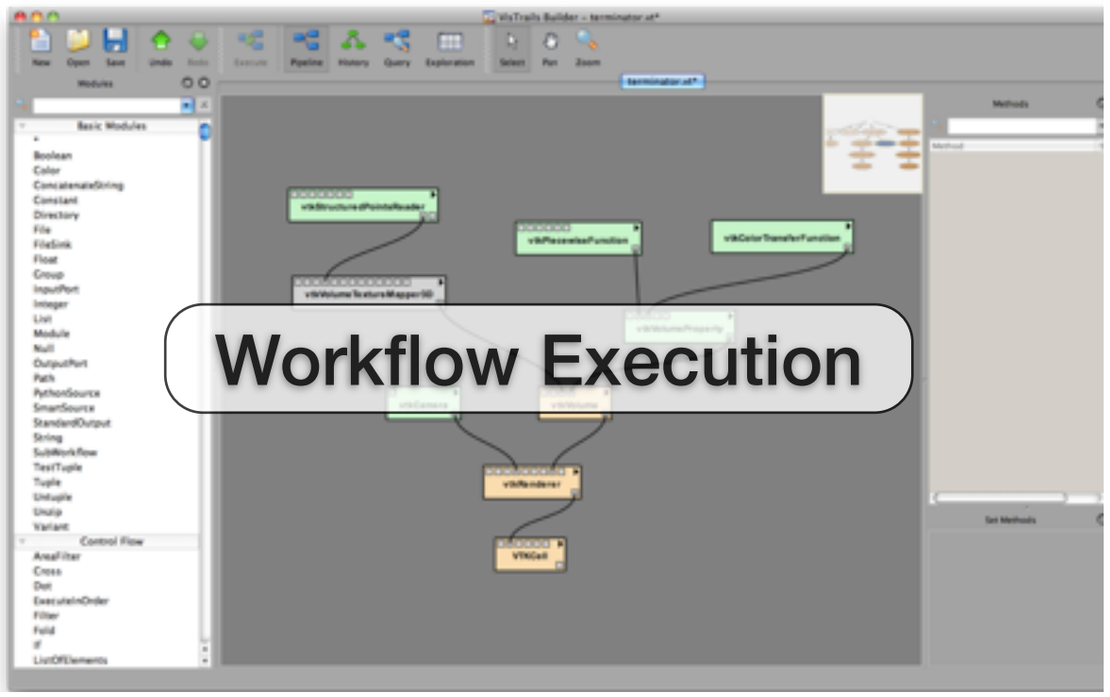
```
identifier = 'org.ocefpaf.seawater'
version = '1.0.3'
name = 'Seawater Routines'

import seawater

class SaturationN2(Module):
    _input_ports = [('S', Float),
                    ('T', Float)]
    _output_ports = [('res', Float)]

    def compute(self):
        s = self.getInputFromPort("S")
        t = self.getInputFromPort("T")
        res = seawater.satN2(s, t)
        self.setResult('res', res)

_modules = [SaturationN2,
```

VisTrails Components





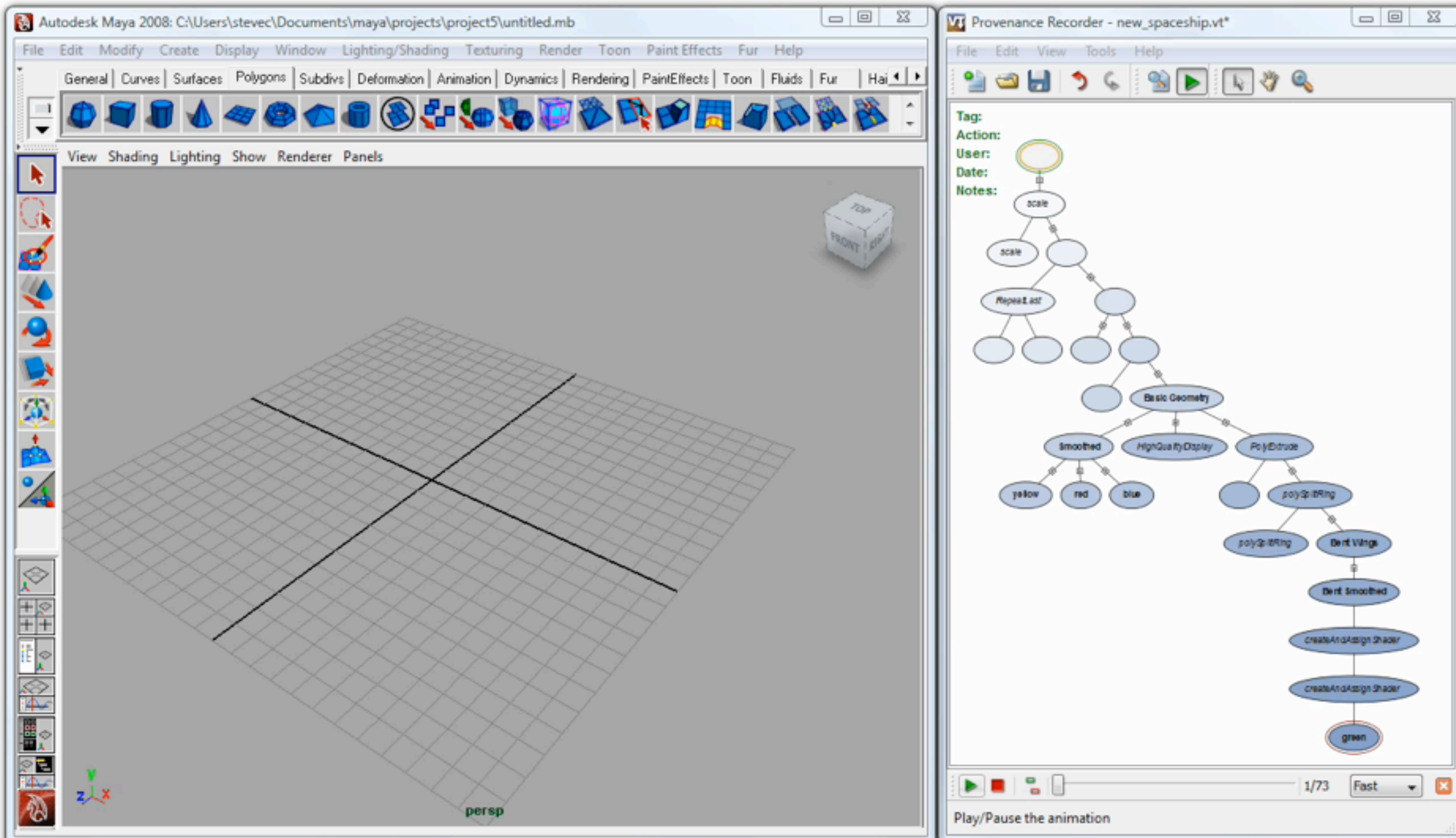




VisTrails Maya Plug-in

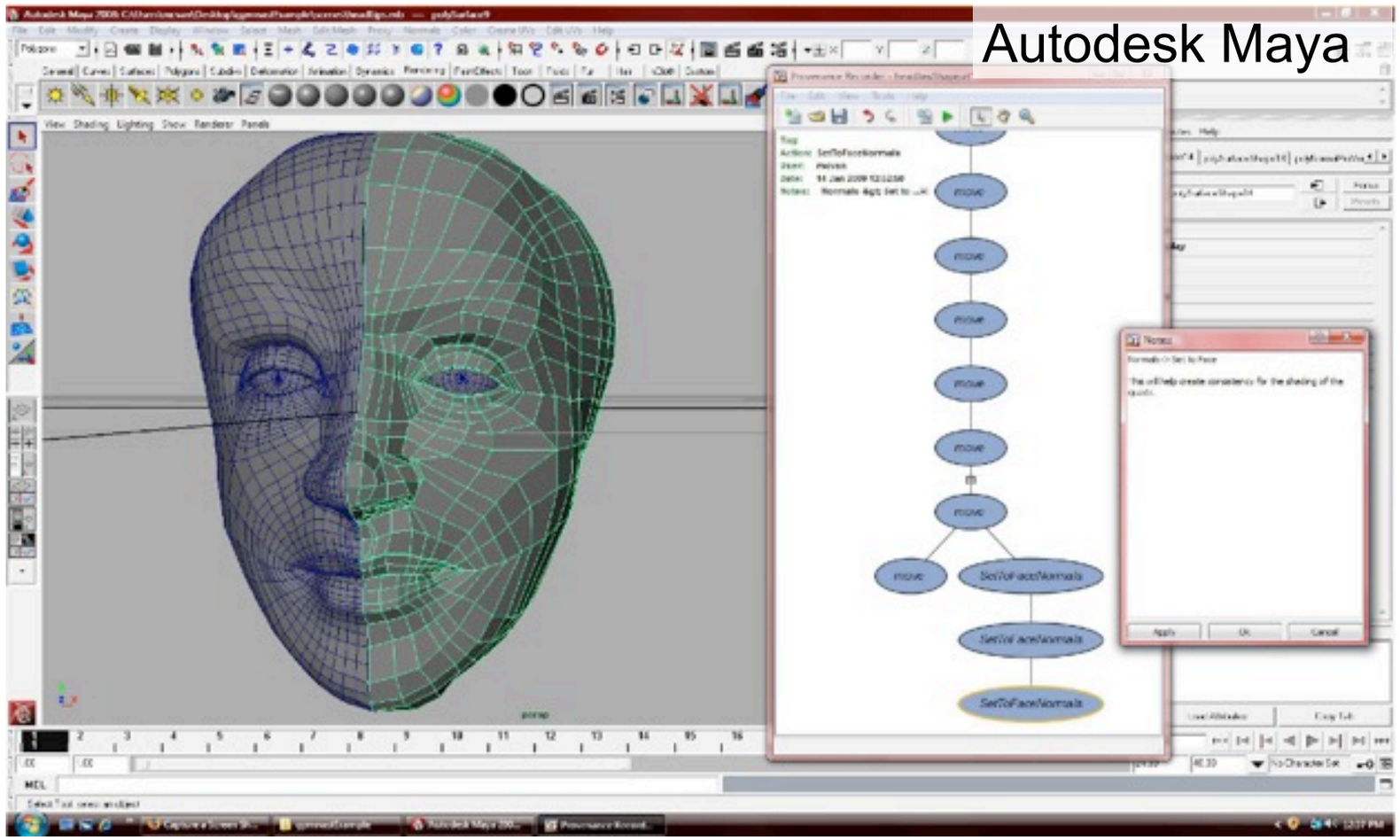
[VisTrails, Inc., IPAW 2008]

VisTrails Maya Plug-in

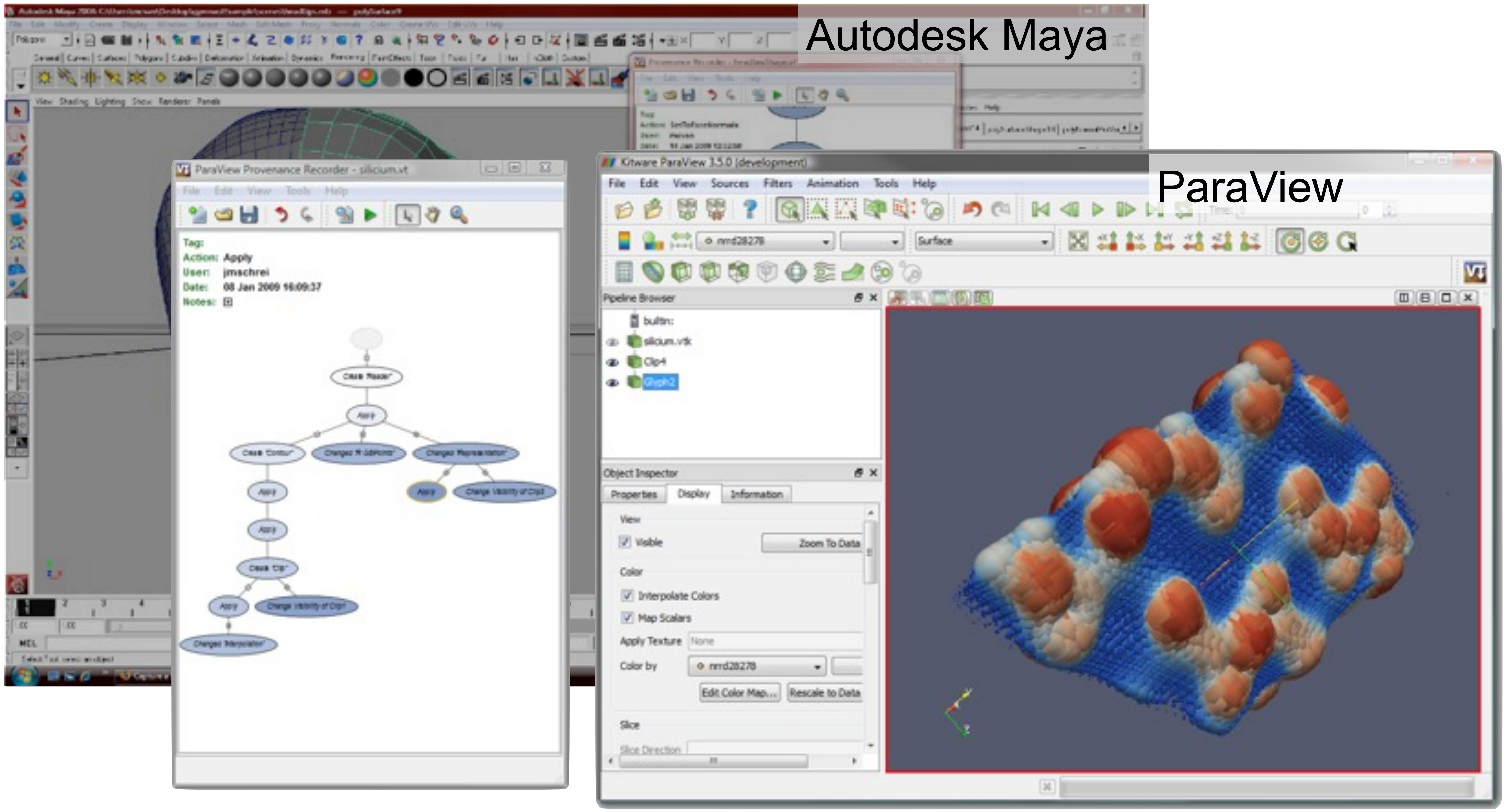


[VisTrails, Inc., IPAW 2008]

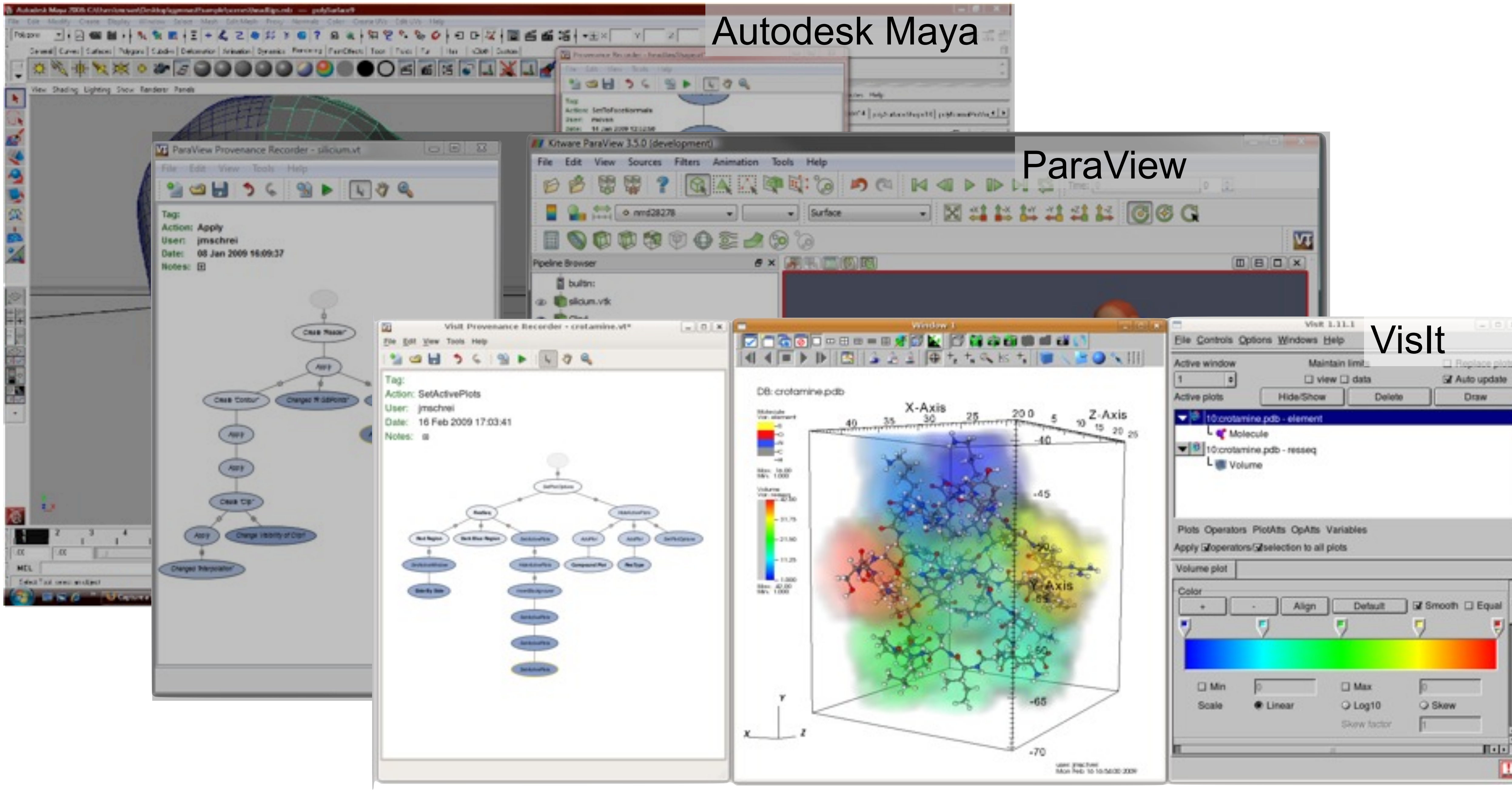
Provenance Enabling 3rd-Party Tools



Provenance Enabling 3rd-Party Tools

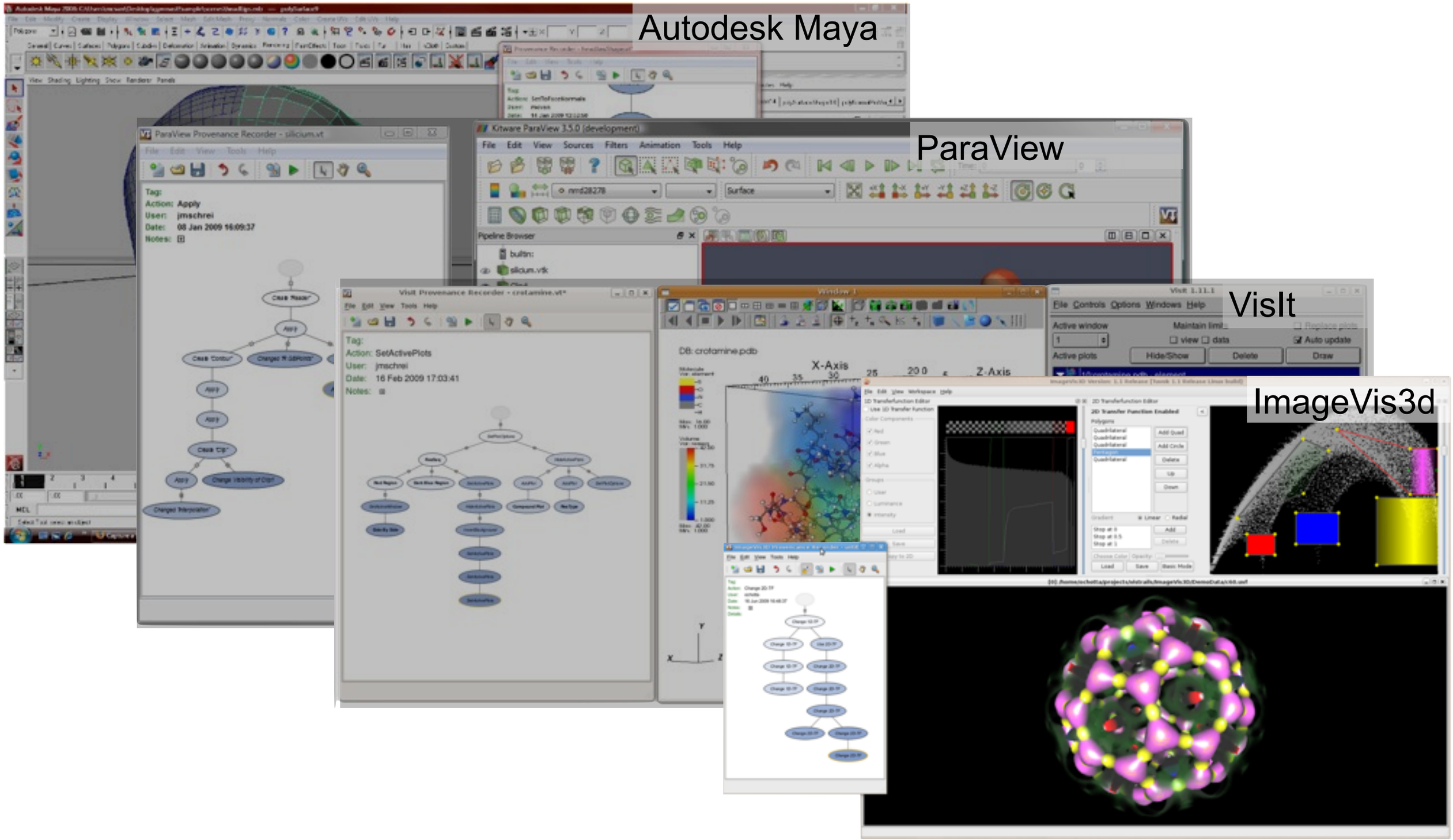


Provenance Enabling 3rd-Party Tools



Friday, February 19, 2010

Provenance Enabling 3rd-Party Tools



Acknowledgments

- **Python Community**
- VisTrails Users & Developers
- VisTrails Team
  - Dr. Cláudio Silva
  - Dr. Juliana Freire

python™

SCI INSTITUTE

THE UNIVERSITY OF UTAH

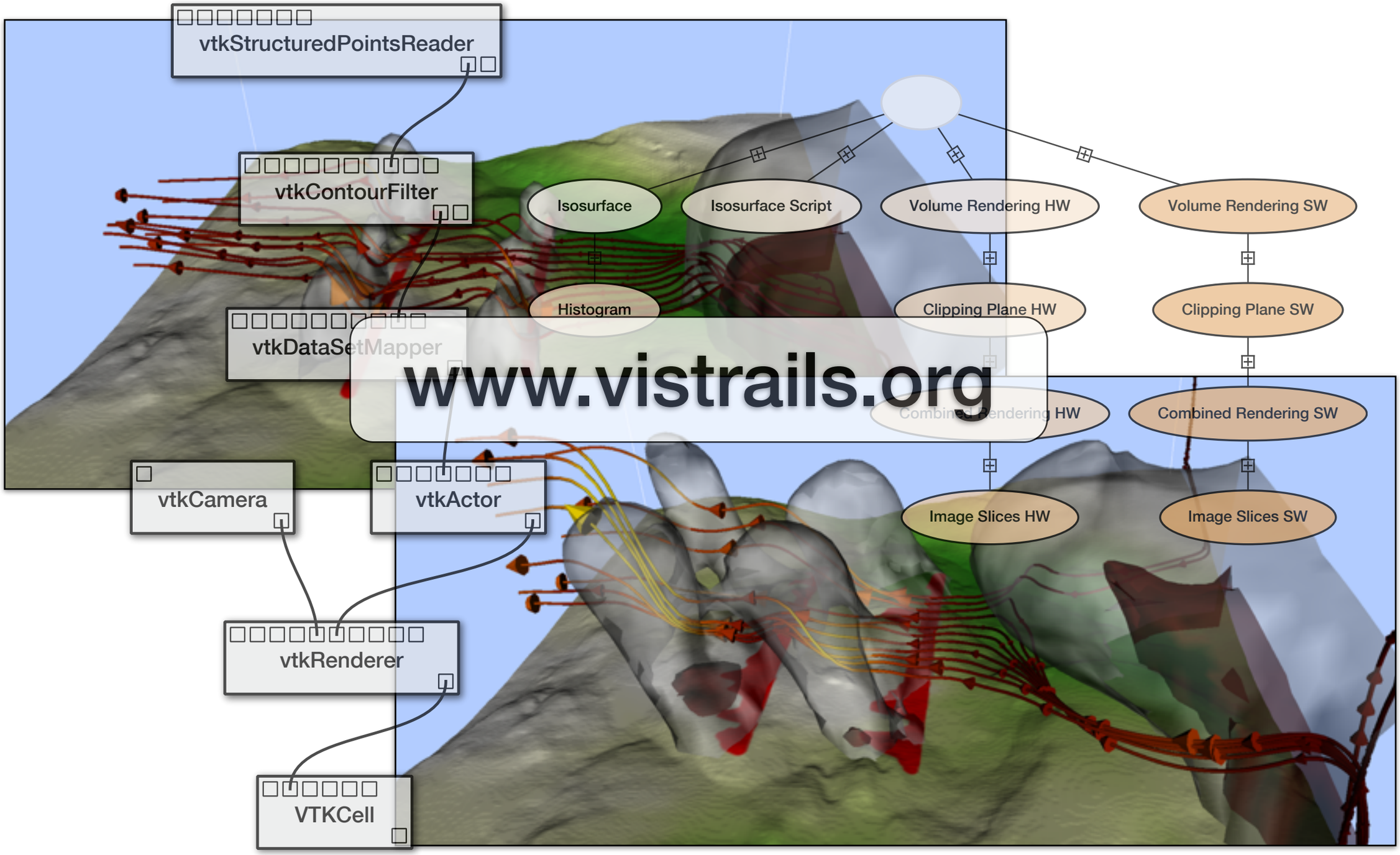
CMOP Center for Coastal Margin Observation & Prediction

U.S. DEPARTMENT OF ENERGY

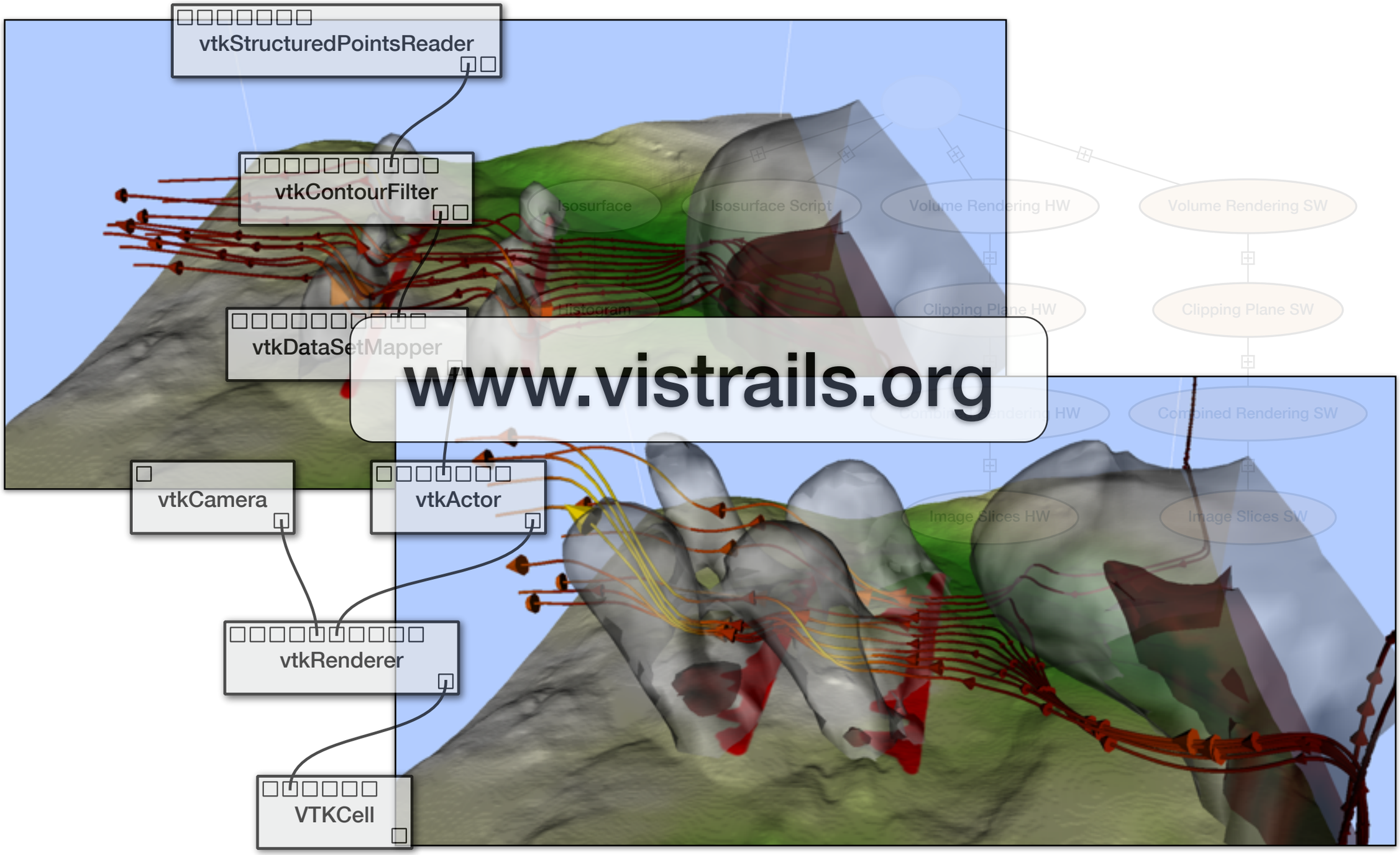
VACET DOE SciDAC VISUALIZATION AND ANALYTICS CENTER FOR ENABLING TECHNOLOGIES

National Science Foundation WHERE DISCOVERIES BEGIN

Questions


vtkStructuredPointsReader
vtkContourFilter
vtkDataSetMapper
vtkCamera
vtkActor
vtkRenderer
VTKCell
Isosurface
Isosurface Script
Volume Rendering HW
Volume Rendering SW
Histogram
Clipping Plane HW
Clipping Plane SW
Combined Rendering HW
Combined Rendering SW
Image Slices HW
Image Slices SW
www.vistrails.org

Questions


vtkStructuredPointsReader
vtkContourFilter
vtkDataSetMapper
vtkCamera
vtkActor
vtkRenderer
VTKCell
www.vistrails.org
Isosurface
Isosurface Script
Volume Rendering HW
Volume Rendering SW
Histogram
Clipping Plane HW
Clipping Plane SW
Combined Rendering SW
Image Slices HW
Image Slices SW